

-1751-

YE TLEK+ +LRAQVA A++PA+AM AS+ LTRQ+S I AVAE+YPLD LKAN +++KLQ
 Sb jct: 61 KYEQATLEKVTQLRAQVASASSPADAMKASDALTRQISGIFAVAESYPLD LKANENYLKLQ 120
 Query: 121 EELTNTENKISYSRQLYN+TTTSSNYNVKLETFPSNIVGKLFQFKPSQFLETPEEEKEVPKV 180
 EELTNTENKISYSRQLYN+ NYNVKL+ FPSN++ +F F+P+ FL TPEEEK VPKV
 Sb jct: 121 EELTNTENKISYSRQLYNSVAGNYNVKQLQAFPSNVIAGMFAFRPADFLSTPEEEKAVPKV 180
 Query: 181 SF 182
 F
 Sb jct: 181 DF 182

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4857> which encodes the amino acid sequence <SEQ ID 4858>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC44350 GB:U66186 LemA [Listeria monocytogenes]
 Identities = 91/181 (50%), Positives = 121/181 (66%), Gaps = 2/181 (1%)
 Query: 5 LIILVVLGV LALWLMISYNSLVKSRMHTKEAWSQIDVQLKRRNDLIPNLIETVKGYASYE 64
 +I + V+ +L L YNSLVK R E W+QIDVQLKRR DLIPNL+ETVKGYA +E
 Sb jct: 5 IIAIAVVVILVLIYFGLYNSLVKYRNRVDETWAQIDVQLKRRFDLIPNLVETVKGYAKHE 64
 Query: 65 QKTFEKITDLRARVAN--ASTPQETMAASNELSKQVTSLFAVAENYPDLKANENFLKLQE 122
 ++T ++ + R ++ A Q + A N LS + S+FA+ E YPLD LKAN +F++LQ
 Sb jct: 65 KETLTQVIEARNKMEVPADNRQQQIEADNMLSGALKSIFALGEAYPDLKANTSFIELQH 124
 Query: 123 ELTNTENKISYSRQLYNSTTSNYNLQLESFPSNLAGKLFQFKPSEFLQTPEAEKEVPKVEF 183
 ELT TENK++YSRQLYN+T YN +++S P+NI KL F + L PE E+ PKVEF
 Sb jct: 125 ELTTTENKVAYSRLYNTTVMYNTKQSVPTNIVAKLHNFTERDMLSIPEVERVAPKVEF 185

An alignment of the GAS and GBS proteins is shown below.

Identities = 135/181 (74%), Positives = 165/181 (90%)
 Query: 4 MILIAIIALFVIWLIVAYNSLVSRMHTKESWSQIDVQLKRRNDLIPNLIETVKGYAAYE 63
 +I++ ++ + +WL+++YNSLV+SRMHTKE+WSQIDVQLKRRNDLIPNLIETVKGYA+YE
 Sb jct: 5 LIILVVLGV LALWLMISYNSLVKSRMHTKEAWSQIDVQLKRRNDLIPNLIETVKGYASYE 64
 Query: 64 GKTLEKIAELRAQVAKANTPAEAMTASNELTRQLSSILAVAENYPDLKANNSFVKLQEEL 123
 KT EKI +LRA+VA A+TP E M ASNEL++Q++S+ AVAENYPDLKAN +F+KLQEEL
 Sb jct: 65 QKTFEKITDLRARVANASTPQETMAASNELSKQVTSLFAVAENYPDLKANENFLKLQEEL 124
 Query: 124 TTNENKISYSRQLYN+TTTSSNYNVKLETFPSNIVGKLFQFKPSQFLETPEEEKEVPKVSFDF 184
 TTNENKISYSRQLYN+TTSNYN++LE+FPSNI GKLFQFKPS+FL+TPE EKEVPKV F+F
 Sb jct: 125 TTNENKISYSRQLYNSTTSNYNLQLESFPSNLAGKLFQFKPSEFLQTPEAEKEVPKVEFNF 185

A related GBS gene <SEQ ID 8849> and protein <SEQ ID 8850> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0
 McG: Discrim Score: 14.63
 GvH: Signal Score (-7.5): -3.19
 Possible site: 20

>>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 1 value: -15.44 threshold: 0.0
 INTEGRAL Likelihood =-15.44 Transmembrane 4 - 20 (1 - 27)

-1752-

PERIPHERAL Likelihood = 8.86 146
modified ALOM score: 3.59

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.7177(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

51.4/68.9% over 183aa

Listeria monocytogenes

EGAD|149857| LemA protein Insert characterized

GP|1519287|gb|AAC44350.1||U66186 LemA Insert characterized

ORF01545(301 - 846 of 1152)

EGAD|149857|159923(2 - 185 of 185) LemA protein {*Listeria monocytogenes*}

GP|1519287|gb|AAC44350.1||U66186 LemA {*Listeria monocytogenes*}

%Match = 23.8

%Identity = 51.4 %Similarity = 68.9

Matches = 94 Mismatches = 56 Conservative Sub.s = 32

42 72 102 132 162 192 222 252
CFK*TSLSVIAVRLIFSFSHSTRSLK*VSNCFCLSVSVIPCSIRT**NAWGIVVNLNFYIV**LYFITNTNNGNNRTFL

282 312 342 372 402 432 462 492
I*RKLL*WKKCKGATTMGTMILIAIIALFVIWLIVAYNSLVSRMHTKESWSQIDVQLKRRNDLIPNLIETVKGYYAAYEG

:| :| ||:: ::|: ||||: | :|:||||| ||||:||||| :|
MIGWIIAIAVVVILVLIYFGLYNSLVKYRNRVDETWAQIDVQLKRRFDLIPNLVETVKGYYAKHEK
10 20 30 40 50 60

522 546 576 606 636 666 696 726
KTLEKIAELRAQVAK--ANTPAEAMTASNELTRQLSSILAVAENYPDLKANNSEFVKLQEELTNTENKISYSRQLYNTTTS

:|| ::| :|: :| :| :| :|:| :| ||||| :|:| ||| ||||:|||||
ETLTQVIEARNKMMVEVPADNRQGQIEADNMLSGALKSIFALGEAYPDLKANTSFIEHQELTTTENKVAYSRLYNTTVM
80 90 100 110 120 130 140

756 786 816 846 876 906 936 966
NYNVKLETFPNSNIVGKLFQFKPSQFLETPEEEKEVPKVSFDF*LRRERGFCCINKLVIREKQLSC*LSSSVF*QLLEQL

|| :|:| :| ||| ||| :| :| :|
TYNTKVQSVPTINIVAKLHNFTERDMLSIPEVERVAPKVEF
160 170 180

SEQ ID 4856 (GBS42) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 2; MW 21.8kDa) and in Figure 168 (lane 5-7; MW 36kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 8; MW 46kDa). Purified Thio-GBS42-His is shown in Figure 244, lane 11.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1572

A DNA sequence (GBSx1666) was identified in *S.agalactiae* <SEQ ID 4859> which encodes the amino acid sequence <SEQ ID 4860>. This protein is predicted to be glucose inhibited division protein b (gidB). Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1753-

bacterial cytoplasm --- Certainty=0.2430(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 10079> which encodes amino acid sequence <SEQ ID 10080> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB16137 GB:Z99124 glucose-inhibited division protein [Bacillus subtilis]
 Identities = 130/239 (54%), Positives = 170/239 (70%), Gaps = 4/239 (1%)

10 Query: 5 MTPQAFYQVLIHGGITLTDKQKKQFETYFRLLEVWNEKINLTAITDKEEVYLKHFYDSIA 64
 M + F L E GI+L+ +Q +QFE Y+ +LVEWNEKINLT+IT+K+EVYLKHFYDSI
 Sbjct: 1 MNIEEFTSGLAEKGISLSRQLEQFELYDMLVEWNEKINLT SITEKKEVYLKHFYDSIT 60

15 Query: 65 PILQGYID-NSPLSILDIGAGAGFSPIMKILYPEIDITIIDSINLRINFLNLANELEL 123
 Y+D N +I D+GAGAGFPS+P+KI +P + +TI+DSLNRKI FL L+ L+L
 Sbjct: 61 AAF--YVDFNQNTICDVGAGAGFPSLPIKICFPHLHVTTIVDSLNRKITFLEKLSEALQL 118

20 Query: 124 SGVHFFHGRAEDFGQDRVFRKFDIVTARAVAKMQVLAELTIPFLKVNRLIALKAAAAE 183
 F H RAE FGQ + R +DIVTARAVA++ VL+EL +P +K NG +ALKAA+AE
 Sbjct: 119 ENTTFCHDRAETFGQRKDVRESYDIVTARAVARLSVLSELCLPLVKNGLFVALKAAASAE 178

Query: 184 EELISAEKALKTLFSQVTVNKNYKLP-NGDDRNTIVSKKETPNKYPRKAGTPNKKPL 241
 EEL + +KA+ TL ++ ++KLP DRNI ++ K K TP KYPRK GTPNK P+
 25 Sbjct: 179 EELNAGKKAITTLGGELNHSFKLPIESDRNIMVIRKIKNTPKKYPRKPGTPNKSPI 237

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4861> which encodes the amino acid sequence <SEQ ID 4862>. Analysis of this protein sequence reveals the following:

Possible site: 21
 30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4862(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 170/237 (71%), Positives = 202/237 (84%)

40 Query: 5 MTPQAFYQVLIHGGITLTDKQKKQFETYFRLLEVWNEKINLTAITDKEEVYLKHFYDSIA 64
 MTPQ FY+ L E G +L+ KQK+QF+TYF+ LVEWN KINLTAIT++ EVYLKHFYDSIA
 Sbjct: 1 MTPQDFYRTLEEDGFSLSKQKEQFDYFKSLVEWNTKINLTAITEENEVYLKHFYDSIA 60

45 Query: 65 PILQGYIDNSPLSILDIGAGAGFSPIMKILYPEIDITIIDSINLRINFLNLANELELS 124
 PILQG++ N P+ +LDIGAGAGFPS+PMKIL+P +++TIIDSINLRKI+FL +LA EL L
 Sbjct: 61 PILQGFLANEPIKLLDIGAGAGFPSLPMKILFPNLEVTTIIDSINLRISFLTLLAQELGLE 120

Query: 125 GVHFFHGRAEDFGQDRVFRKFDIVTARAVAKMQVLAELTIPFLKVNRLIALKAAAAEE 184
 VHFFHGRAEDFGQD+ FR +FD+VTARAVA+MQVL+ELTIPFLK+ G+LIALKA AA++
 50 Sbjct: 121 NVHFFHGRAEDFGQDKAFRGQFDVVVTARAVARMQVLSLTTIPFLKIGGKLIALKAQAADQ 180

Query: 185 ELISAEKALKTLFSQVTVNKNYKLPNGDDRNTIVSKKETPNKYPRKAGTPNKKPL 241
 EL A+ AL LF +V N +Y+LPNGD R ITIV KKKETPNKYPRKAG PNKKPL
 55 Sbjct: 181 ELEEAKNALCLLFGKVIKNHSYQLPNGDSRFITIVKKKETPNKYPRKAGLPNKKPL 237

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1754-

Example 1573

A DNA sequence (GBSx1667) was identified in *S.agalactiae* <SEQ ID 4863> which encodes the amino acid sequence <SEQ ID 4864>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1574

A DNA sequence (GBSx1668) was identified in *S.agalactiae* <SEQ ID 4865> which encodes the amino acid sequence <SEQ ID 4866>. This protein is predicted to be v-type sodium ATP synthase subunit j. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -10.14 Transmembrane 371 - 387 (362 - 391)

INTEGRAL Likelihood = -7.48 Transmembrane 200 - 216 (190 - 217)

INTEGRAL Likelihood = -4.94 Transmembrane 425 - 441 (423 - 446)

INTEGRAL Likelihood = -4.67 Transmembrane 327 - 343 (325 - 349)

INTEGRAL Likelihood = -3.77 Transmembrane 81 - 97 (81 - 98)

INTEGRAL Likelihood = -2.66 Transmembrane 140 - 156 (139 - 157)

INTEGRAL Likelihood = -1.33 Transmembrane 55 - 71 (53 - 71)

INTEGRAL Likelihood = -0.27 Transmembrane 247 - 263 (247 - 263)

INTEGRAL Likelihood = -0.11 Transmembrane 165 - 181 (165 - 181)

----- Final Results -----

bacterial membrane --- Certainty=0.5055(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10081> which encodes amino acid sequence <SEQ ID 10082> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA04279 GB:D17462 Na+ -ATPase subunit J [Enterococcus hirae]

Identities = 170/461 (36%), Positives = 262/461 (55%), Gaps = 28/461 (6%)

Query: 12 KTMSVARKLSISFIAVILLGSLLSLPIFQYANAPKTHYIDHLFTTVSMVCVTGLSVFPPI 71

K +S + ++ F +IL G LL+LP F + TH+ID LFT S VCVTGL+

Sbjct: 10 KRLSPVQLIAAGFFILILFGGSLTLPPFFS-RSGESTHFIDALFTATSACVCTGLTTLNT 68

Query: 72 SKVYNGWQIVAILLMQTGGLGLVITLMSLSYYTLRRKMSLNDQTLLQSAITYNSSTDLLKK 131

++ +N GQ + + L++ GGLG + + L + ++K+S + + +L+ A+ + + K

Sbjct: 69 AEHWNSAGQFLIMTLIEIGGLGFMIPILFFAIAKKKISFSMRIVLKEALNLEEMSGVIK 128

Query: 132 YLYMIFKVTLTLEVLAAASILAIIDFIPRFGLGHGIFNSIFLAVSAFCNAGFDNLEATSLAQ 191

+ I K + ++V+ A L++ FIP FG GI+ SIF AVS+FCNAGFD L + LA

Sbjct: 129 LMIYILKFAVVIQVIGAVALS VVFIPEFGWAKGIWFSIFHAVSSFCNAGFDLLGDSLLAD 188

-1755-

Query: 192 FKLNPLVNIIVCFLLIISGGLGFAVWKDLIEATIQTSKGPGLIKTFPKRLSNHSLVLKT 251
 + N + ++V LII+GGLGF VW+D++ + H+ K+++ HSK+ L
 Sbjct: 189 -QTNVYLIMVVSALIIAGGLGFIVWRDIL-----SYHR-----VKKITLHSHKVALSV 234

5 Query: 252 TTIILLTGTLSSWLEFGNFRITIANLSLPKQLMVSFFQTVMTAGFSTIDYTQTDFAFN 311
 T ++L+ G +L +L+ N T+ + ++L +FF +VT RTAG+ +IDY Q A
 Sbjct: 235 TALLLIGGFIL-FLITERNGLTLVKGTFTTERLANTFFMSVTPRTAGYYSIDYLQMSHAGL 293

10 Query: 312 LVYIIQMLIGGAPGGTAGGFKVTVIAILLLLFKAELSGSQVTFHYRTIPSSIQKTL SI 371
 ++ + M IGG G TAGG K T + ILL+ A G+++ RTI + + L
 Sbjct: 294 IITMFLMYIGGTSGSTAGGLKTTTLGILLIQHAMFKGKTRAEAFGRTIRQAAV---LRA 350

15 Query: 372 LTFFFI--LISGYLLLELNPIDPFS----LFFEASSALATVGVIMNTTNQLTLGGRI 425
 LT FF+ L +++L + I S + FE SA TVG+TM T LTL G++
 Sbjct: 351 LTLFFVTLSLCVVAIMVLSVTETIPKTSIEYIAFEVFSAFGTVGLTMGLTPDLTLIGKL 410

20 Query: 426 VIMFLMFIGRVGPITVLLSILQK---KEKEIHYAETEILG 463
 VI+ LM+IGRVG +TV+LS+L K E Y E I+LG
 Sbjct: 411 VIISLMYIGRVGIMTVLSSLVKANRAEANYKYPEESIMLG 451

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4867> which encodes the amino acid sequence <SEQ ID 4868>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -15.12	Transmembrane	371 - 387 (364 - 396)
INTEGRAL	Likelihood = -7.32	Transmembrane	20 - 36 (18 - 42)
INTEGRAL	Likelihood = -6.53	Transmembrane	425 - 441 (417 - 446)
INTEGRAL	Likelihood = -6.16	Transmembrane	89 - 105 (81 - 106)
INTEGRAL	Likelihood = -5.79	Transmembrane	200 - 216 (196 - 223)
INTEGRAL	Likelihood = -3.35	Transmembrane	140 - 156 (139 - 157)
INTEGRAL	Likelihood = -3.03	Transmembrane	55 - 71 (53 - 74)
INTEGRAL	Likelihood = -3.03	Transmembrane	247 - 263 (246 - 264)
INTEGRAL	Likelihood = -1.12	Transmembrane	393 - 409 (393 - 409)
INTEGRAL	Likelihood = -0.11	Transmembrane	165 - 181 (165 - 181)

----- Final Results -----

bacterial membrane --- Certainty=0.7050(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAA04279 GB:D17462 Na+ -ATPase subunit J [Enterococcus hirae]
 Identities = 168/466 (36%), Positives = 260/466 (55%), Gaps = 26/466 (5%)

45 Query: 6 MKRSFIKSLSVTQRLTFSFAIVILIGTLLSMPFTHYQNGPNTVYLDHFFNVVSMVCVTG 65
 MK+ K LS Q + F I+IL G LL++PF ++G +T ++D F S VCVTG
 Sbjct: 4 MKKRVRKRLSPVQLIAAGFFLILFGGSLTLTPFFS-RSGESTHFIDALFTATSACVVTG 62

50 Query: 66 LSVVPVAEVYNGIGQTIAMALMQIGCLGLVTLIAVSTFAL-KRKMRLSDQTLQSAALNRG 124
 L+ + AE +N GQ + M L++IG LG + +I + FA+ K+K+ S + +L+ ALN
 Sbjct: 63 LTTLNTAEHWNASAGQLIMTLIEIGGLGFM-MIPILFFAIKKKISFSMRIVLKEALNLE 121

55 Query: 125 DSKDLKHYLFFAYKVTFSLFAFAIVIMIDFIPRFGWKNGIFNSIFLAVSAFCNAGFDNL 184
 + + + + K ++ A+ + + FIP FGW GI+ SIF AVS+FCNAGFD L
 Sbjct: 122 EMSGVIKLMIYILKFAVVIQVIGAVLSVVFPEFGWAKGIWFSIFHAVSSFCNAGFDLL 181

60 Query: 185 GSSSLKDFMLNPTLNVIITFLIISGGLGFAVWVDLGVAFKKYFFERPHCYGATFRKLSNQ 244
 G S L D N L ++++ LII+GGLGF VW D+ +++ + +K++
 Sbjct: 182 GDSLLAD-QTNVYLIMVVSALIIAGGLGFIVWRDI-LSYHR-----VKKITLH 227

65 Query: 245 SRLVLQTTAVILFLGTFLTWFLFKDNTIANFSLHQQLMVSFFQTVIMRTAGFATISYN 304
 S++ L TA++L +G F+ + + + N T+ + ++L +FF +VT RTAG+ +I Y
 Sbjct: 228 SKVALSVTALLL-IGGFILFLITERNGLTLVKGTFTTERLANTFFMSVTPRTAGYYSIDYL 286

Query: 305 DTLAPTNIYMIQMVIGGAPGGTAGGKIVTTAAITFLLLFKAELSGQSEVTFNRRIANKT 364

-1756-

IL M M IGG G TAGG+K TT I + A G++ R I
 Sbjct: 287 QMSHAGLLTLMFLMYIGGTSGSTAGGLKTTTLGILLIQMHAMFKGKTRAEAFGRTIRQAA 346
 Query: 365 IKQITMTVLIFFFFAVLMIIGFILLLSVEPHIAPIP----LLFESISAIATVGVSMDLTPQLS 420
 + + +T L F L + I++LSV I + FE SA TVG++M LTP L+
 Sbjct: 347 VLRALT-LFFVTLSLCVVAIMVLSVTETIPKTSIGIEYIAFEVFSAFGTVGLTMGLTPDLT 405
 Query: 421 TAGRLIVIVLMFVGRVGPITVLSLI---QRKEKTIQYATTDILVG 463
 G+L++I LM++GRVG +TV++SL+ R E +Y I++G
 Sbjct: 406 LIGKLVIISLMYIGRVGIMTVVLSLLVKANRAEANYKYPEESIMLG 451

An alignment of the GAS and GBS proteins is shown below.

Identities = 275/462 (59%), Positives = 351/462 (75%), Gaps = 1/462 (0%)

Query: 2 GASMKHFDDYKTMISVARKLSISFIAVILLGSILLSLPFQYANAPKTHYIDHLFTTVMV 61
 G +MK F K++SV ++L+ SF VIL+G++LLS+P Y N P T Y+DH F VSMV
 Sbjct: 3 GGNMKRSF-IKSLSVTQRLTFSAIVILIGTLLSMPFTHYQNGPNTVYLDHFFNVVSMV 61
 Query: 62 CVTGLSVFPISKVYNGWGQIVAILLMQTGGLGLVTLMSLSYYTLRRKMSLNDQTLQSAI 121
 CVTGLSV P+++VYNG GQ +A+ LMQ G LGLVTL+++S + L+RKM L+DQTLQSA+
 Sbjct: 62 CVTGLSVVPVAEVYNGIGQTIAMALMQIGCLGLVTLIAVSTFALKRKMRLSDQTLQSAI 121
 Query: 122 TYNSTDLKKLYMIFKVTLTLEVLAAISILAIIDFIPRFGLGHGIFNSIFLAVSAFCNAGF 181
 S DLK YL+ +KVT +LE AA ++ IDFIPRFG +GIFNSIFLAVSAFCNAGF
 Sbjct: 122 NRGDSKDLKHLYLFAYKVTFSLEAFAAIVIMIDFIPRFGWKNIGIFNSIFLAVSAFCNAGF 181
 Query: 182 DNLEATSLAQFKLNPLVNIIVCFIISGGLGFAVWVDLIEATIQTSHKGPKLIKTFPKRL 241
 DNL ++SL F LNP +N+I+ FLIISGGLGFAVW DL A + + P ++L
 Sbjct: 182 DNLGSSSLKDFMLNPTLNVIITFLIISGGLGFAVWVDLGVAFKKYFFERPHCYGATFRKL 241
 Query: 242 SNHSLVLKTTTIIILLTGTLLSWLLEFGNFRTIANLSLPKQLMVSFFQVTMRTAGFSTI 301
 SN S+LVL+TT +IL GT L+W LE N +TIAN SL +QLMVSFFQVTMRTAGF+TI
 Sbjct: 242 SNQSRLVLQTAVILFLGTFTLWFLFKDNSKTIANFSLHQQLMVSFFQVTMRTAGFATI 301
 Query: 302 DYTQTFATNLVYIIQMLIGGAPGGTAGGKVTVIAILLLLFKAELSGQSQVTFHYRTIP 361
 Y T TN++Y+IQM+IGGAPGGTAGG KVT AI LLFKAELSGQS+VTF R I
 Sbjct: 302 SYNDTLAPTNIYMIQMVIGGAPGGTAGGKVTAAITFLFKAELSGQSEVTFRNRIIA 361
 Query: 362 SSIKQTLISILTFFFIILISGYLLLELNPIDPFSLFFEASSALATVGVMTNTNQLTL 421
 + IKQT++L FFF +L+ G++LLL + PHI P L FE+ SA+ATVGV+M+ T QL+
 Sbjct: 362 NKTIKQTMVLIFFFFAVLMIIGFILLLSVEPHIAPILLFESISAIATVGVSMDLTPQLST 421
 Query: 422 GGRIVIMFLMFIGRVGPITVLLSILQKKEKEIHYAETEILG 463
 GR++++ LMF+GRVGPITVL+S++Q+KEK I YA T+I++G
 Sbjct: 422 AGRILIVIVLMFVGRVGPITVLSLIQRKEKTIQYATTDILVG 463

A related GBS gene <SEQ ID 8851> and protein <SEQ ID 8852> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9
 McG: Discrim Score: 0.86
 GvH: Signal Score (-7.5): 0.64
 Possible site: 45
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 9 value: -10.14 threshold: 0.0
 INTEGRAL Likelihood = -10.14 Transmembrane 371 - 387 (362 - 391)
 INTEGRAL Likelihood = -7.48 Transmembrane 200 - 216 (190 - 217)
 INTEGRAL Likelihood = -4.94 Transmembrane 425 - 441 (423 - 446)
 INTEGRAL Likelihood = -4.67 Transmembrane 327 - 343 (325 - 349)
 INTEGRAL Likelihood = -3.77 Transmembrane 81 - 97 (81 - 98)
 INTEGRAL Likelihood = -2.66 Transmembrane 140 - 156 (139 - 157)
 INTEGRAL Likelihood = -1.33 Transmembrane 55 - 71 (53 - 71)
 INTEGRAL Likelihood = -0.27 Transmembrane 247 - 263 (247 - 263)
 INTEGRAL Likelihood = -0.11 Transmembrane 165 - 181 (165 - 181)
 PERIPHERAL Likelihood = 2.49 308
 modified ALOM score: 2.53

*** Reasoning Step: 3

```
5      bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

```

10      ORF02334(334 - 1689 of 1989)
      EGAD|22151|22827(10 - 451 of 451) v-type sodium ATP synthase subunit j {Enterococcus hirae}
      SP|P43440|NTPJ_ENTHR V-TYPE SODIUM ATP SYNTHASE SUBUNIT J (EC 3.6.1.34) (NA(+)-
      TRANSLOCATING ATPASE SUBUNIT J). GP|487282|dbj|BAA04279.1||D17462 Na+ -ATPase subunit J
      {Enterococcus hirae}
15      %Match = 18.8
      %Identity = 38.5 %Similarity = 60.4
      Matches = 170 Mismatches = 166 Conservative Sub.s = 97

```

25 426 456 486 516 546 576 606 636
 ILLSLPIQYANAPKTHYIDHLFTTVSMVCVTGLSVFPISKVYNGWGQIVAILLMQTGGGLGLVTLMSLSYTTLRRKMSLN
 ||:||| : ||:||| | ||||| : : : : || : : ||| : : | : : :
 SLLTLPPFSRSGES-THFIDALFTATSAVCVTGLTTLNTLAEHWSNAGQFLIMTLIEIGGLGFMMPIPLFFAIAKKISFS
 40 50 60 70 80 90 100

[illegible]

906 936 966 996 1026 1056 1086 1116
KLNPLVNIIVCFLIIISGGLGFAVWKDLIEATIQTSHKGPKLIKTFFPKRLSNHSKLVLKTITTIILLTGILLSWLLEFGNFR
:
| : | : ||||| : | :
QTNVYLIMVSALIIAGGLGFIVWRDI-----LSYHRVKKITLHSKVALSVTA-LLIGGFILFLITERNGL

200 210 220 230 240 250

45

1146	1176	1206	1236	1266	1296	1326	1356
TIANLSLPKQLMVSFFQIVTMRTAGFSTIDYTDYQDFATNLVYIIQMLIGGAPGGTAGGFKVTVIAILLLLFKAELSGQSQ							
: :::: : : : : ::: : : : :::							
TLVKGTFTERLANTFFMSVTPRTAGGYISIDYQLQMSHAGLILTMFLMYIGTSGSTAGGLKTTTLGILLIQMHAMFKGKTR							
270		280		290		300	
310		320		330			

50 1386 1416 1461 1491 1518 1548 1578
VTFHYRTIPSSIIKQTLISILTFFFIIL---ISGYLL-LLELNPHIDPFS-LFFEASSALATVGVTMTNTNQLTLGGRIV
||| : : | ||::| : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
AEAAGRTIRQAAV--LRALTLFVLTSLCVAIMVLVSVTETIPKTSIEIYAEFVFSAFGTVGLTMGLTPDLTLIGKLIV

350 360 370 380 390 400 410

```

55      1608      1638      1659      1689      1719      1749      1779      1809
IMPLMFIGRVGPITVLLSILQK---KEKEIHYAETETIILG*KRSFMKTKIIGVLGLGIFGQTLAQELSNEFQDVIAIDSN
|: ||: ||| | :||: ||: | | | | | :||
IISLMYIGRVGIMTVVLSLLVKANRAEANYKYPEESIMLG
              430              440              450

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1758-

Example 1575

A DNA sequence (GBSx1669) was identified in *S.agalactiae* <SEQ ID 4869> which encodes the amino acid sequence <SEQ ID 4870>. This protein is predicted to be TrkA (ktrA). Analysis of this protein sequence reveals the following:

5 Possible site: 19
 >>> Seems to have a cleavable N-term signal seq.

 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 10 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAC46144 GB:AF001974 putative TrkA [Thermoanaerobacter
 ethanolicus]
 Identities = 69/177 (38%), Positives = 110/177 (61%), Gaps = 2/177 (1%)

Query: 8 VLGLGIFGQTLAQELSNFEQDVIAIDSNPEN--VQAAEVVTKAAIGDITDLAFLKHIGI 65
 V+GLG FG +LA+ L DV+ ID + E VQA+ +VT A D TD LK + +
 20 Sbjct: 6 VIGLGSFGISLAKTLYEMGNDVLVIDEEDDEEELVQAMNGLVTHAVRADATDENVLKSLRV 65

Query: 66 SDCDTVIIATGNSLESSVLAVMHCKKLGVPQVIKARNLVYEEVLYEIGADLVISPERES 125
 + D I+A G ++ESS++ M K+LGV VIAKA N ++ VLY++GAD V+ PE++
 25 Sbjct: 66 KNFDVAIVAIGKNMESSIMVTMLVKELGVKYVIAKAHNEHARVLYKVGADRVVMPEKDM 125

Query: 126 GQNVAAANLMRNKITDVFQIESDISVIEFKIPKSWVGKTVEQLNIRHKFDLNLIGIRK 182
 G VA N+ + + D+ + + S+ E + W GKT++++N+R K+ LN++ ++K
 30 Sbjct: 126 GIRVARNVFSSNLIDLIEFSKEYSIAEILPIEEWFGKTLKEINVREKYGLNVVAVKK 182

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4715> which encodes the amino acid sequence <SEQ ID 4716>. Analysis of this protein sequence reveals the following:

 Possible site: 20
 >>> Seems to have an uncleavable N-term signal seq

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

 Identities = 132/221 (59%), Positives = 176/221 (78%)

Query: 1 MKTKIIGVLGLGIFGQTLAQELSNFEQDVIAIDSNPENVQAAEVVTKAAIGDITDLAFL 60
 +K K +GVLGLGIFG+T+A+ELSNF+QDVIAID +V+ VA++VTKAA+GDITD FL
 45 Sbjct: 2 LKRKTGVLGLGIFGRTVARELSNFDQDVIAIDIRESHVKEVADLVTKAAVGDITDKEFL 61

Query: 61 KHIGISDCDTVIIATGNSLESSVLAVMHCKKLGVPQVIKARNLVYEEVLYEIGADLVIS 120
 +GI CDTV+IA+GN+LESSVLAVMHCKKLGVP +IAKA+N ++EEVLY IGA VI+
 50 Sbjct: 62 LAVGIEHCDTVVIASGNNLESSVLAVMHCKKLGVPITIIAKAKNKIFEVLYGIGATKVIT 121

Query: 121 PERESGQNVAAANLMRNKITDVFQIESDISVIEFKIPKSWVGKTVEQLNIRHKFDLNLIGI 180
 PER+SG+ VA+NL+R I + +E IS+IEF IPKSW G+++ +L++R K++LN+IG+
 55 Sbjct: 122 PERDSGKRVASNLLRRHIESIITYLEHGISMIEFVIPKSWEGQSLSELDVRRKYELNVIGM 181

Query: 181 RKAKNKPVDTEVPINSPLEEGIILVAIANSDAFQRYDYLG 221
 R+ + K +DT V PLE I+VAIAN F+++DYLG
 Sbjct: 182 RQKEVKTLDTNVKPFEPLEPNTIIVAIAANDHTFEKFDYLG 222

-1759-

A related GBS gene <SEQ ID 8853> and protein <SEQ ID 8854> were also identified. Analysis of this protein sequence reveals the following:

```

5  Lipop: Possible site: -1  Crend: 3
   McG: Discrim Score:      5.14
   GvH: Signal Score (-7.5): -0.860001
       Possible site: 19
   >>> Seems to have a cleavable N-term signal seq.
   ALOM program  count: 0 value: 1.06 threshold: 0.0
10  PERIPHERAL Likelihood = 1.06 192
   modified ALOM score: -0.71

   *** Reasoning Step: 3

   ----- Final Results -----
15  bacterial outside --- Certainty=0.3000(Affirmative) < succ>
     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
     bacterial cytoplasm --- Certainty=0.0000(Not Clear)

```

The protein has homology with the following sequences in the databases:

```

20  38.0/61.6% over 182aa
                                     Thermoanaerobacter
   ethanolicus
   GP|2581796| putative TrkA Insert characterized

25  ORF02030(322 - 864 of 1269)
   GP|2581796|gb|AAC46144.1|AF001974(6 - 188 of 195) putative TrkA {Thermoanaerobacter
   ethanolicus}
   %Match = 15.5
   %Identity = 37.9 %Similarity = 61.5
30  Matches = 69 Mismatches = 69 Conservative Sub.s = 43

   60      90      120      150      180      210      240      270
   LISGYLLLELNP HIDPFS LFFEASSALATVGVTMNTTNQLTLGGRIVIMFLMFIGRVGPITVLLSILQKKEKEIHYAET

35  300      330      360      390      444      474      504
   EIILG*KRSFMKTKIIGVLGLGIFGQTLAQELSNFEQDVIAIDSNPEN--VQVAEVVTKAAIGDITDLAFLKHIGISDC
       |  |||  ||  ||:  |  ||:  ||  :  |  |||:  :||  |  ||  ||  :  :
       MKQFVVIGLGSFGISLAKTLYEMGNDVLVIDEEDDEELVQAMNGLVTHAVRADATDENVLKSLRVKNF
           10      20      30      40      50      60

40  534      564      594      624      654      684      714      744
   DTVIIATGNSLESSVLAVMHCKKLGVPQVIKARNLVKEEVLYEIGADLVISPERESGQNVAAANLMRNKITDVFQIESDI
   |  :|  |  :|||:  |  |||  |||  |  :  |||:  ||  |  ||  :  :  :  :
   DVAIVAIGKNMESSIMVMTLVKELGVKYVIAKAHNEHARVLYKVGADRVVMPEKDMGIRVARNVFSSNLIDLIEFSKEY

45  80      90      100      110      120      130      140

   774      804      834      864      894      924      954      984
   SVIEFKIPKSWVGKTVEQLNIRHKFDLNLIGIRKAKNKPVDTEVPINSPLEEXIILVAIANSDAFQRYDYLRIFY*RK*K
   |  |  :  |  |||:  :||  |  ||:  :|  :  :  :
50  SIAEILPIEWFQKTLKEINVREKYGLNVVAVKKFNDDEIIVSPGAGL
       160      170      180      190

```

SEQ ID 8854 (GBS57) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 6; MW 26kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 11; MW 51.1kDa) and in Figure 183 (lane 9 & 10; MW 51kDa).

The GBS57-GST fusion product was purified (Figure 99A; see also Figure 195, lane 8) and used to immunise mice (lane 1 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 99B), FACS (Figure 99C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

-1760-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1576

A DNA sequence (GBSx1670) was identified in *S.agalactiae* <SEQ ID 4871> which encodes the amino acid sequence <SEQ ID 4872>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.62	Transmembrane	73 - 89 (68 - 96)
INTEGRAL	Likelihood = -11.30	Transmembrane	254 - 270 (248 - 274)
INTEGRAL	Likelihood = -4.73	Transmembrane	127 - 143 (124 - 144)
INTEGRAL	Likelihood = -4.19	Transmembrane	50 - 66 (47 - 67)
INTEGRAL	Likelihood = -3.29	Transmembrane	25 - 41 (25 - 45)

----- Final Results -----

bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8855> which encodes amino acid sequence <SEQ ID 8856> was also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 9

McG: Discrim Score: -10.49

GvH: Signal Score (-7.5): -1.14

Possible site: 40

>>> Seems to have no N-terminal signal sequence

ALOM program count: 5 value: -11.62 threshold: 0.0

INTEGRAL	Likelihood = -11.62	Transmembrane	73 - 89 (68 - 96)
INTEGRAL	Likelihood = -11.30	Transmembrane	254 - 270 (248 - 274)
INTEGRAL	Likelihood = -4.73	Transmembrane	127 - 143 (124 - 144)
INTEGRAL	Likelihood = -4.19	Transmembrane	50 - 66 (47 - 67)
PERIPHERAL	Likelihood = 3.76		201

modified ALOM score: 2.82

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13178 GB:Z99110 ykoC [Bacillus subtilis]

Identities = 61/226 (26%), Positives = 108/226 (46%), Gaps = 12/226 (5%)

Query: 49 FLIVVSLGSLVLFRLAKIKWQVSVFMTLVVVFVAVLNIIIMVYLFAPHYGDKIYGSSSLLL 108
F I++ G L+ + KW + + F +L V+ A K+ + L
Sbjct: 36 FYIIIVAGVLLAAGIPLKKW-----LLFTIPFLILAFGCVWTAAVF--GKVPTTPDNFL 87

Query: 109 KGIGPYDVTSQELFYLFNLIKYFCTVPLALLFLMTTNPSQFASSL-NQLGLSYKIAYAV 167
GP + S + +L + C L+++F+ TT+P F SL Q LS K+AY V
Sbjct: 88 FQAGPISINSDNVSVGISLGFRLCFSSALSMFVFTTDPILFMLSILVQQCRLSPKLAYGV 147

Query: 168 SLTLRYIPDVQEEFYTIIRRAQEARGIELSKSNLVARIKGNLQIVTPLIFSSLERIDTVA 227
R++P +++E I++A + RG + +S ++ +I + PL+ S++ + + A
Sbjct: 148 IAGFRFLPLLKDEVQLIQQAHKIRGG--AAESGIINKISALKRYTIPLLASAIRKAERTA 205

Query: 228 TAMELRRFGKNKRRTWYSKQSLKSDIVLIILALASLFVSLYLIHL 273
AME + F ++ RT+Y S+ + D V L L LF +L+ L
Sbjct: 206 LAMESKGFTGSRNRTYYRTLSVNRRDWVFFCLVLL-LFAGSFLVSL 250

-1761-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1577

A DNA sequence (GBSx1671) was identified in *S.agalactiae* <SEQ ID 4873> which encodes the amino acid sequence <SEQ ID 4874>. This protein is predicted to be cobalt ABC transporter, ATP-binding protein (cbiO). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.91    Transmembrane  436 - 452 ( 435 - 452)

----- Final Results -----
          bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13179 GB:Z99110 similar to cation ABC transporter
(ATP-binding protein) [Bacillus subtilis]
Identities = 151/483 (31%), Positives = 248/483 (51%), Gaps = 19/483 (3%)

Query: 8   KDFTFYQYDVQSEPTLKGINLSIPKGEKVLILGPSGSGKSTLGHCNLNGIIPNTHKGQYSGI 67
+   +F Y+   +P + I+   + KGE VL+LGPSG GKS+L CLNG+ P   G   SG
Sbjct: 11  EQLSFSYEEDEKPVFQDISFELQKGEVLLLGPSGCGKSSIALCLNGLYPEACDGIQSGH 70

Query: 68  FTINHKNADFSLSIYDK-SHLVSTVLQDPDQFIGLTVAEDIAFALENDVVAQEEMASIVE 126
+   K   D   +   +   V QDPD QF LTV ++IAF LEN + +EEM +
Sbjct: 71  VFLFQKPVTDAAETSETITQHAGVVFDQDPDQFCMLTVEDEIAFGLLENLQIPKEEMTEKIN 130

Query: 127 MWAKRLEIAPLLSKRPQDLSSGGQKQVSLAGVLVDDSPILLFDEPLANLDPQSGQDIMAL 186
+L I L K   LSGGQKQ+V+LA +L + +++ DEP + LDP S ++ + L
Sbjct: 131 AVLGLKRITHLKEKMISTLSGGQKQKVALACILAMEPELIILDEPTSLLDPFSAREFVHL 190

Query: 187 VDRIHQEQDATTTIIIEHRLED--VFYERVDRVVLFSDDGQIIYNGEPDQLL--KTNFLSEY 242
+   + +E+   + ++IEH+L++   + ER   +VL   G+   +G   L   +   L +
Sbjct: 191 MKDLQREKGFSLLVIEHQLEWAPWIERT--IVLDKSGKKALDGLTKNLFQHEAETLKKL 248

Query: 243 GIREPLYISALKKNLGYDFEKQNTMTSIDDFFSELLIPKMRALDLKHTDKLLSVQHLSV 302
GI P   + L F   M   +   + K +A   + +L V LS
Sbjct: 249 GIAIPKVCHLQEKLSMPFTLSKEMLFKEPIAGH--VKKKKA---PSGESVLEVSSLSF 302

Query: 303 SYDLENNITLDDVSFDLYKGQRLAIVGKNGAGKSTLAKALCQFI-PNNATLIYNNEDVSQD 361
+   +   D+SF L +G   A+VG NG GKSTL L + P + ++ ++ + +
Sbjct: 303 ARG-QQAIFKDISFSLREGSLTALVGPNGTGKSTLLSVLASLMKPQSGKILLYDQPLQKY 361

Query: 362 SIKERAERIGYVLQNPQMISQAMVFDEVALGLRLRGFSDNDIESRVYDILKVCGLYQFR 421
KE +R+G+V QNP   V+DE+ G +   ++ + E +   +L+ GL
Sbjct: 362 KEKELRKRMGFVFQNPQHVFVTDTVYDELLFGQK---ANAETEKKAQHLLQRFGLAHLA 417

Query: 422 NWPISALSFGQKKRVTTIASILINPEVIILDEPTAGQDMKHYTEMMSFLDKLSCDGHITIV 481
+   A+S GQK+R+++A++L+ + +V++LDEPT GQD +   E M + ++ +G ++
Sbjct: 418 DHHPFAISQGGKRRLSVATMLMHDVKVLLLEPTFGQDARTAAECMEMIQRKAEGTAVL 477

Query: 482 MIT 484
MIT
Sbjct: 478 MIT 480
```

There is also homology to SEQ ID 4416.

SEQ ID 4874 (GBS424d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 2 & 4; MW 77kDa) and in Figure 239 (lane 10; MW 77kDa). It

-1762-

was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 5 & 7; MW 52kDa) and in Figure 182 (lane 4; MW 52kDa). Purified GBS424d-His is shown in Figure 241, lanes 6 & 7. Purified GBS424d-GST is shown in Figure 246, lane 12.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1578

A DNA sequence (GBSx1672) was identified in *S.agalactiae* <SEQ ID 4875> which encodes the amino acid sequence <SEQ ID 4876>. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.12	Transmembrane	39 - 55 (35 - 63)
INTEGRAL	Likelihood = -3.98	Transmembrane	72 - 88 (71 - 90)
INTEGRAL	Likelihood = -3.66	Transmembrane	108 - 124 (106 - 127)
INTEGRAL	Likelihood = -2.34	Transmembrane	182 - 198 (181 - 198)
INTEGRAL	Likelihood = -1.44	Transmembrane	141 - 157 (139 - 158)

----- Final Results -----

bacterial membrane	---	Certainty=0.4248(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB59830 GB:AJ012388 hypothetical protein [Lactococcus lactis]

Identities = 109/182 (59%), Positives = 141/182 (76%)

Query: 31	MNTNTIKKVATGIGAALEFIIIGMLVNIPTPIPTNIQLQYAVLALFAVIYGPVGFFTG	90
	M N++K VVATGIGAALEF+IIG L+NIPTPIPTNT+IQLQYAVLALF+ ++GP GF G	
Sbjct: 1	MKNNSVKIVVATGIGAALEFVIIGWLINIPTPIPTNTSIQLQYAVLALFSALEFGFLAGFLIG	60
Query: 91	FIGHALKDSIQYGSPWWTWVLVSGLLGLMIGFFAKKLAIQLSGMTKKDLLLLFNVVQVIAN	150
	FIGHALKDS YG+PWWTWVL SGL+GL +GF K+ ++ K+++ FN+VQ +AN	
Sbjct: 61	FIGHALKDSFLYGAPWWTWVLGSGLMGLFLGFGVKRESLTQGIFGNKEIIRFNIVQFLAN	120
Query: 151	LIGWSVVPYGDIFFYSEPASKVFQAQGLSSLVNSITIGVGGTLLLLLAYAKSRPQKGSLS	210
	++ W ++AP GDI YSEPA+KVF QG ++ LVN++TI V GTLLL YA +R + G+L	
Sbjct: 121	VVWGLIAPIGDILVYSEPANKVFTQGVVAGLVNALTIAVAGTLLLLKYAATRTKSGTLD	180

Query: 211 KD 212

K+

Sbjct: 181 KE 182

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8857> and protein <SEQ ID 8858> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6

McG: Discrim Score: -5.01

GvH: Signal Score (-7.5): -5.9

Possible site: 50

>>> Seems to have no N-terminal signal sequence

ALOM program count: 5 value: -8.12 threshold: 0.0

INTEGRAL	Likelihood = -8.12	Transmembrane	31 - 47 (27 - 55)
INTEGRAL	Likelihood = -3.98	Transmembrane	64 - 80 (63 - 82)
INTEGRAL	Likelihood = -3.66	Transmembrane	100 - 116 (98 - 119)
INTEGRAL	Likelihood = -2.34	Transmembrane	174 - 190 (173 - 190)
INTEGRAL	Likelihood = -1.44	Transmembrane	133 - 149 (131 - 150)
PERIPHERAL	Likelihood = 5.78		9

modified ALOM score: 2.12

-1764-

Example 1580

A DNA sequence (GBSx1674) was identified in *S.agalactiae* <SEQ ID 4879> which encodes the amino acid sequence <SEQ ID 4880>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -3.61	Transmembrane	107 - 123 (96 - 124)
INTEGRAL	Likelihood = -1.86	Transmembrane	124 - 140 (124 - 142)
INTEGRAL	Likelihood = -1.38	Transmembrane	83 - 99 (83 - 100)
INTEGRAL	Likelihood = -1.12	Transmembrane	142 - 158 (142 - 160)

----- Final Results -----

bacterial membrane	---	Certainty=0.2444 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9415> which encodes amino acid sequence <SEQ ID 9416> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC76124 GB:AE000391 putative transport protein [Escherichia coli K12]
Identities = 139/178 (78%), Positives = 159/178 (89%)

Query: 1 MVGTMFLFVALVVPNPIAFVMMRKNPYPLVLRCLKDSGITAFFTRSSAANIPVNMRLCEDL 60
+VG ML VALVVPNP++ + +R+NP+PLVL CL++SG+ AFFTRSSAANIPVNM LCE L
Sbjct: 222 LVGCMLLVALVVPNPLLVWVKIRRNPPPLVLLCLRESGVYAFFTRSSAANIPVNMALCEKL 281

Query: 61 GLDKDTYSVSISPLGAAINMAGAAITINILTLAAVNTLGIITVDFTAFLLSVVAAVSACGA 120
LD+DTYSVSISPLGA INMAGAAITI +LTLAAVNTLGI VD PTA LLSVVA++ ACGA
Sbjct: 282 NLDRTYSVSISPLGATINMAGAAITITVLTLLAAVNTLGIIPVDLPTALLSVVASLCACGA 341

Query: 121 SGVTGGSLLLIPLVACSLFGISNDVAMQVVGVGFIQDSCETALNSSTDVLFATA 178
SGV GGSLLLIPLVACSLFGISNDVAMQVVGFI+GV+QDSCETALNSSTDVLFATA A
Sbjct: 342 SGVAGGSLLLIPLACNMFGISNDIAMQVAVGVFIIGVLQDSCETALNSSTDVLFATAA 399

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4881> which encodes the amino acid sequence <SEQ ID 4882>. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -13.69	Transmembrane	212 - 228 (202 - 239)
INTEGRAL	Likelihood = -7.38	Transmembrane	78 - 94 (74 - 108)
INTEGRAL	Likelihood = -6.53	Transmembrane	179 - 195 (175 - 200)
INTEGRAL	Likelihood = -6.10	Transmembrane	315 - 331 (312 - 341)
INTEGRAL	Likelihood = -5.36	Transmembrane	44 - 60 (42 - 61)
INTEGRAL	Likelihood = -4.41	Transmembrane	13 - 29 (11 - 41)
INTEGRAL	Likelihood = -3.19	Transmembrane	340 - 356 (333 - 358)
INTEGRAL	Likelihood = -3.08	Transmembrane	145 - 161 (144 - 162)
INTEGRAL	Likelihood = -0.90	Transmembrane	358 - 374 (358 - 376)

----- Final Results -----

bacterial membrane	---	Certainty=0.6477 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

>GP:AAF95950 GB:AE004347 sodium/dicarboxylate symporter [Vibrio cholerae]
Identities = 243/385 (63%), Positives = 299/385 (77%), Gaps = 2/385 (0%)

Query: 9 VRVSLIKIGIGVIGVMLGILAPDLTG-FSILGKLFVGGGLKAIAPLLVFALVSQAISHQ 67
VR +L+ +I G+++G + +P+ ++G LFVG LKA+AP+LVF LV+ +I++Q

-1765-

5 Sbjct: 11 VRGNLVLQILAGILLGAAMATFSPEYAQKVGLIGNLFVGALKAVAPVLVLFILVASSIANQ 70
 Query: 68 KKGKQTNMTLIIIVLYLFGTFASALVAULTAYLFPITLVNTPVNTLSPPQGVAEVFQSL 127
 KK + T M I+VLYLFGTF++AL AV+ ++LFP TLVL T +PPQG+AEV +L
 Sbjct: 71 KKNQHTYMRPIVVLYLFGTFSAALTAVILSFLFPTTLVLATGAEGA-TPPQGI AEVLNLT 129
 10 Query: 128 LLKLVDNPINALATANYIGVLSWAIIFGLLKAASKETKHLIKTAAEVTSQIVVWIINLA 187
 L KLVDNP++AL ANYIG+L+W + GLAL +S TK + + + SQIV +II LA
 Sbjct: 130 LFKLVDNPVSALMNANYIGILAWGVGLGLALHSSSTTKAVFEDLSHGISQIVRFIIRLA 189
 Query: 188 PIGIMSLVFTTISENGVGILSDYAFLILVLVGTMLFVALVNNPLIAVLITRQNPYPLVLR 247
 P GI LV +T + G L+ YA L+ VL+G M F+ALVNNP+I R+NP+PLVL+
 Sbjct: 190 PFGIFGLVASTFATTGFDALAGYAQLLAVLLGAMAFIALVNNPMIVYYKIRRNPFPLVLQ 249
 15 Query: 248 CLRESGLTAFTRSSAANIPVNMQLCQKIGLSKDTYSVSIPLGATINMGAAITINVLTL 307
 CLRESG+TAFTRSSAANIPVNM LC+K+ L +DTYSVSIPLGATINM GAAITI VLTL
 Sbjct: 250 CLRESGVTAFTRSSAANIPVNMALCEKLDKEDTYSVSIPLGATINMAGAAITITVLTL 309
 20 Query: 308 AAVHTFGIPIDFLTALLLSVVAASACGASGVAGGSLLLIPVACSLFGISNDLAMQVVG 367
 AAVHT GI +D +TALLSVVAASACGASGVAGGSLLLIP+AC LFGISND+AMQV V
 Sbjct: 310 AAVHTMGIEVDLMTALLLSVVAASACGASGVAGGSLLLIPLACGLFGISNDIAMQVAV 369
 Query: 368 GFIVGVQDSCETALNSSTDVLF 392
 GFI+GVQDS ETALNSSTDVLF
 25 Sbjct: 370 GFIVGVQDSCETALNSSTDVLF 394

An alignment of the GAS and GBS proteins is shown below.

Identities = 153/186 (82%), Positives = 172/186 (92%)

30 Query: 1 MVGTMLFVALVNNPIIAFVMMRKNPYPLVLRCLKDSGITAFFTRSSAANIPVNMRLCEDL 60
 +VGTMFLFVALVNNP+IA ++ R+NPYPLVLRCL++SG+TAFTRSSAANIPVNM+LC+ +
 Sbjct: 217 LVGTMLFVALVNNPLIAVLITRQNPYPLVLRCLRESGLTAFTRSSAANIPVNMQLCQKI 276
 35 Query: 61 GLDKDTYSVSIPLGAANMAGAAITINILTLAAVNTLGITVDFPTAFLLSVVAASACGA 120
 GL KDTYSVSIPLGA INM GAAITIN+LTLAAV+T GI +DF TA LLSVVAASACGA
 Sbjct: 277 GLSKDTYSVSIPLGATINMGAAITINVLTLAAVHTFGIPIDFLTALLLSVVAASACGA 336
 Query: 121 SGVTGGSLLIPVACSLFGISNDVAMQVVGFGFIVGVQDSCETALNSSTDVLF 180
 SGV GGSLLIPVACSLFGISND+AMQVVGFGFIVGVQDSCETALNSSTDVLF+AE
 40 Sbjct: 337 SGVAGGSLLLIPVACSLFGISNDLAMQVVGFGFIVGVQDSCETALNSSTDVLF+AE 396
 Query: 181 SVFGKK 186
 + + +K
 45 Sbjct: 397 AFWKRK 402

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1581

50 A DNA sequence (GBSx1675) was identified in *S.agalactiae* <SEQ ID 4883> which encodes the amino acid sequence <SEQ ID 4884>. This protein is predicted to be acid phosphatase. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2436(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1766-

A related GBS nucleic acid sequence <SEQ ID 9427> which encodes amino acid sequence <SEQ ID 9428> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:CAA73175 GB:Y12602 acid phosphatase [Streptococcus equisimilis]
    Identities = 167/251 (66%), Positives = 209/251 (82%)

    Query: 7  EQKTKFKNISLSSNKLAKENTMSVLWYQNSAEAKALYLQGYNVAKMKLDDWLQKPSEKP 66
          ++ K  ++ S  +L +  ENIMSVLWYQ +AEAKALYLQGY +A  +L + L + ++KP
    Sbjct: 34  KETVKQTKVTSDEQLRSNENTMSVLWYQRAAEAKALYLQGYQLATDRLKNQLGQATDKP 93

10  Query: 67  YSIILDLDDETVDNSPYQAKNIKDSSFTPEWDKWKVQKSAKAVAGAKEFLKYANEKGI 126
          YSI+LD+DETVDNSPYQAKNI +G+SFTPEWD WVQKK AK VAGAKEFL++A++ G+
    Sbjct: 94  YSIVLDIDETVDNSPYQAKNILEGTSFTPEWDVWVQKKEAKPVAGAKEFLQFADQNGV 153

15  Query: 127 KIYYVSDRTDAQVDATKENLEKEGIPVQGDHLLFLKKGMSKESRRQAVQKDTNLIIMLF 186
          +IYY+SDR +QVDAT ENL+KEGIPVQG+DHLLFL++G+KSKE+RRQ V++ TNLIIMLF
    Sbjct: 154 QIYYISDRAVSQVDATMENLQKEGIPVQGRDHLLFLEEGVKSKEARRQKVKETTNLIIMLF 213

20  Query: 187 GDNLVDFADFSKSSSTDREQLLTKLQSEFGSKFIVFPNPMYGSWESAIYQGHLDVQKQL 246
          GDNLVDFADFSK S DR LL++LQ EFG +FI+FPNPMYGSWESA+Y+G LD QL
    Sbjct: 214 GDNLVDFADFSKSEEDRTALLSELQEEFGRFIIFPNPMYGSWESAVYKGDKLDASHQL 273

    Query: 247 KERQKMLHSYD 257
          KER+K L S++
25  Sbjct: 274 KERRKALESFE 284

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4885> which encodes the amino acid sequence <SEQ ID 4886>. Analysis of this protein sequence reveals the following:

```

30  Possible site: 25

    >>> May be a lipoprotein

    ----- Final Results -----
35  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

40  >GP:CAA73175 GB:Y12602 acid phosphatase [Streptococcus equisimilis]
    Identities = 234/284 (82%), Positives = 261/284 (91%)

    Query: 1  MKSKKVSVISLTLFLVLTGCAKVDNNKSVNLKPATKQTYNSYSDQLRSRENTMSVLW 60
          MK+K+V SVISL LSLFLVTGCA++D+ +VN K KQT +YSD+QLRS ENIMSVLW
    Sbjct: 1  MKTKQVASVISLALSLFLVTGCAQLDHKANVNSKETVVKQTKVTSDEQLRSNENTMSVLW 60

45  Query: 61  YQRAAEQALYLQGYQLATDRLKEQLNKPTDKPYSIVLDIDETVDNSPYQAKNVLEGTG 120
          YQRAAE +ALYLQGYQLATDRLK QL + TDKPYSIVLDIDETVDNSPYQAKN+LEGT
    Sbjct: 61  YQRAAEAKALYLQGYQLATDRLKNQLGQATDKPYSIVLDIDETVDNSPYQAKNILEGTS 120

50  Query: 121 FTPESWDYVWQKKEAKPVAGAKDFLQFADQNGVQIYYISDRSTTQVDATMENLQKEGIPV 180
          FTPESWD WVQKKEAKPVAGAK+FLQFADQNGVQIYYISDR+ +QVDATMENLQKEGIPV
    Sbjct: 121 FTPESWDVWVQKKEAKPVAGAKEFLQFADQNGVQIYYISDRAVSQVDATMENLQKEGIPV 180

55  Query: 181 QGRDHLLFLEKGVKSKESSRRQKVKETTNVTMLFGDNLDFADFSKKSQEDRTALLSDLQE 240
          QGRDHLLFLE+GVKSKE+RRQKVKETTN+ MLFGDNL+DFADFSKKS+EDRTALLS+LQE
    Sbjct: 181 QGRDHLLFLEEGVKSKEARRQKVKETTNLIIMLFGDNLVDFADFSKKSSEEDRTALLSELQE 240

    Query: 241 EFGRRFIIFPNPMYGSWEGAIYKGEKLDVLKQLEERRKSLKSKF 284
          EFGR+FIIFPNPMYGSWE A+YKG+KLD QL+ERRK+L+SF+
60  Sbjct: 241 EFGRQFIIFPNPMYGSWESAVYKGDKLDASHQLKERRKALESFE 284

```

An alignment of the GAS and GBS proteins is shown below.

-1767-

Identities = 166/247 (67%), Positives = 207/247 (83%)

Query: 10 TKFKNISLSSNKLAKENTMSVLWYQNSAEAKALYLQGYNVAKMKLDLQKPSEKPYSI 69
 TK S S ++L ++ENTMSVLWYQ +AE +ALYLQGY +A +L + L KP++KPYSI
 5 Sbjct: 37 TKQTYNSYSDDQLRSRENTMSVLWYQRAAETQALYLQGYQLATDRLEQLNKPTDKPYSI 96

Query: 70 ILDLDETFLDNSPYQAKNIKDGSSFTPESEWDKQKSAKAVAGAKEFLKYANEKGIKIY 129
 +LD+DETFLDNSPYQAKN+ +G+ FTPESEWD WVQKK AK VAGAK+FL++A++ G++IY
 10 Sbjct: 97 VLDIDETFLDNSPYQAKNVLEGTGFTPESEWDYVWQKKEAKPVAGAKDFLQFADQNGVQIY 156

Query: 130 YVSDRTDAQVDATKENLEKEGIPVQGDHLLFLKKGMSKESRRQAVQKDTNLIMLFQDN 189
 Y+SDR+ QVDAT ENL+KEGIPVQG+DHLLFL+KG+KSKESESRQ V++ TN+ MLFQDN
 15 Sbjct: 157 YISDRSTQVDATMENLQKEGIPVQGRDHLLFLEKGVKSKESESRQVKETTNVTMLFQDN 216

Query: 190 LVDFAFDFSKSSSTDREQLLTKLQSEFGSKFIVFPNPMYGSWESAIYQGHLDVQKQLKER 249
 L+DFADFDFSK S DR LL+ LQ EFG +FI+FPNPMYGSWE AIY+G+ LDV KQL+ER
 20 Sbjct: 217 LLDFAFDFSKSQEDRTALLSDLQEEFGRRFIIFPNPMYGSWEGAIYKGEKLDVLKQLEER 276

Query: 250 QKMLHSY 256
 +K L S+
 20 Sbjct: 277 RKSLKSF 283

SEQ ID 9428 (GBS661) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell
 extract is shown in Figure 136 (lane 2 & 4; MW 61kDa + lane 3; MW 27kDa) and in Figure 186 (lane 11;
 25 MW 61kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
 extract is shown in Figure 136 (lane 5-7; MW 25kDa).

GBS661-GST was purified as shown in Figure 237, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

30 Example 1582

A DNA sequence (GBSx1676) was identified in *S.agalactiae* <SEQ ID 4887> which encodes the amino
 acid sequence <SEQ ID 4888>. This protein is predicted to be unnamed protein product. Analysis of this
 protein sequence reveals the following:

Possible site: 58
 35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3462(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4889> which encodes the amino acid
 sequence <SEQ ID 4890>. Analysis of this protein sequence reveals the following:

Possible site: 58
 45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3462(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 395/398 (99%), Positives = 398/398 (99%)

55 Query: 1 MAKLTVKDVLKGGKVLVRVDFNVPLKDGVI'TDNRITAALPTIKYIIEQGGRAILFSHL 60

-1768-

MAKLTVKDNDLKGKKVLVRVDFNVPLKDGVTNDNRITAALPTIKYIEQGGRAILFSLH
 Sbjct: 1 MAKLTVKDNDLKGKKVLVRVDFNVPLKDGVTNDNRITAALPTIKYIEQGGRAILFSLH 60
 Query: 61 GRVKEEADKEGKSLAPVAADLAAKLGQDVVFPVTRGAKLEEAINALEDGQVLLVENTRF 120
 GRVKEEADKEGKSLAPVAADLAAKLGQDVVFPVTRG+KLEEAINALEDGQVLLVENTRF
 Sbjct: 61 GRVKEEADKEGKSLAPVAADLAAKLGQDVVFPVTRGSKLEEAINALEDGQVLLVENTRF 120
 Query: 121 EDVDGKKESKNDDEELGKYWASLGDGIFVNDAFGTAHRAHASNVGISANVEKAVAGFLEN 180
 EDVDGKKESKNDDEELGKYWASLGDGIFVNDAFGTAHRAHASNVGISANVEKAVAGFLEN
 Sbjct: 121 EDVDGKKESKNDDEELGKYWASLGDGIFVNDAFGTAHRAHASNVGISANVEKAVAGFLEN 180
 Query: 181 EIAYIQEAVETPERPFVAILGGSKVSDKIGVIENLLEKADKVLIGGGMTYTFYKAQGIEI 240
 EIAYIQEAVETPERPFVAILGGSKVSDKIGVIENLLEKADKVLIGGGMTYTFYKAQGIEI
 Sbjct: 181 EIAYIQEAVETPERPFVAILGGSKVSDKIGVIENLLEKADKVLIGGGMTYTFYKAQGIEI 240
 Query: 241 GNSLVEEDKLDVAKDLLEKSNGKLILPVDKSKEANAFAGYTEVRDTEGEAVSEGFLGLDIG 300
 GNSLVEEDKLDVAKDLLEKSNGKLILPVDKSKEANAFAGYTEVRDTEGEAVSEGFLGLDIG
 Sbjct: 241 GNSLVEEDKLDVAKDLLEKSNGKLILPVDKSKEANAFAGYTEVRDTEGEAVSEGFLGLDIG 300
 Query: 301 PKSIKFDALTGAKTVVWNGPMGVFENPDFQAGTIGVMDAIVKQPGVKSIIGGDSAAA 360
 PKSIA+FD+ALTGAKTVVWNGPMGVFENPDFQAGTIGVMDAIVKQPGVKSIIGGDSAAA
 Sbjct: 301 PKSIAEFDQALTGAKTVVWNGPMGVFENPDFQAGTIGVMDAIVKQPGVKSIIGGDSAAA 360
 Query: 361 AINLGRADKFSWISTGGGASMELEGGKVLPLGLAALTEK 398
 AINLGRADKFSWISTGGGASMELEGGKVLPLGLAALTEK
 Sbjct: 361 AINLGRADKFSWISTGGGASMELEGGKVLPLGLAALTEK 398

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 1583

A DNA sequence (GBSx1677) was identified in *S.agalactiae* <SEQ ID 4891> which encodes the amino acid sequence <SEQ ID 4892>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.39 Transmembrane 97 - 113 (93 - 118)
 INTEGRAL Likelihood = -3.66 Transmembrane 25 - 41 (24 - 48)
 INTEGRAL Likelihood = -3.40 Transmembrane 121 - 137 (121 - 140)
 INTEGRAL Likelihood = -3.24 Transmembrane 72 - 88 (72 - 88)
 INTEGRAL Likelihood = -2.07 Transmembrane 143 - 159 (143 - 160)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4354(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4893> which encodes the amino acid sequence <SEQ ID 4894>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.23 Transmembrane 97 - 113 (93 - 118)
 INTEGRAL Likelihood = -7.17 Transmembrane 121 - 137 (119 - 140)
 INTEGRAL Likelihood = -4.19 Transmembrane 25 - 41 (24 - 48)
 INTEGRAL Likelihood = -3.24 Transmembrane 72 - 88 (72 - 88)
 INTEGRAL Likelihood = -2.55 Transmembrane 154 - 170 (154 - 170)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4291(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1769-

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```

5      Identities = 155/178 (87%), Positives = 169/178 (94%)

Query: 1  MKTLKKLLSNYKFDIKKFKLGMRTFKTGLSVFLVLLVFHFLFGWKGLQIGALTAVFSLRED 60
      MRTL+KLLSNYKFDIKKFKLGMRT KTGLSVFLVLLVFHFLFGWKGLQIGALTAVFSLRED
Sbjct: 1  MKTLRKLSSNYKFDIKKFKLGMRTLKTGLSVFLVLLVFHFLFGWKGLQIGALTAVFSLRED 60

10  Query: 61  FDKSVHFGFSRIIGNSIGLLSLVFFAFNEIFHQAFWVTLIVPICTMLCIMINVACNNK 120
      FDKSVHFGFSRIIGNSIGLLSLVFFAFNEIFHQAFWVTLIVPICTMLCIM+NVACNNK
Sbjct: 61  FDKSVHFGFSRIIGNSIGLLSLVFFAFNEIFHQAFWVTLIVPICTMLCIMVNVACNNK 120

15  Query: 121 SGIIGGTAALLIITLSIPSGETILYVFARIFETFCGVFIAMMVNTDIEILRKKLKNK 178
      SGIIG  AALLIITLSIP+G+T +YV +R+FETFCGVF+A++VNTD+E+++ K N K
Sbjct: 121 SGIIGAVAALLIITLSIPTGTQTFIYVTSRVFETFCGVFVAILVNTDVELIKNKWFNKK 178

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 1584

A DNA sequence (GBSx1678) was identified in *S.agalactiae* <SEQ ID 4895> which encodes the amino acid sequence <SEQ ID 4896>. This protein is predicted to be regulatory protein glr (glrR). Analysis of this protein sequence reveals the following:

```

25      Possible site: 17
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
30      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA00402 GB:D00513 ORF129 [Bacillus cereus]
      Identities = 59/123 (47%), Positives = 89/123 (71%), Gaps = 5/123 (4%)

35  Query: 4  RELRRTMAVFPIGAVMKLTDLTARQIRYYEDQGLITPERTEGNRRMFSLNDMDRLLEIKD 63
      +E RR+ +FPIG VM LT L+ARQIRYYE+ L++P RT+GNRR+FS ND+D+LLEIKD
Sbjct: 2  KEDRRSAPLFPPIGIVMDLTQLSARQIRYYEEHNLVSPTRTKGNRRLFSFNDVDKLLLEIKD 61

40  Query: 64  FISDGLHISDIKNEYMQRQH----KSKEKQKSLSDAEVRRLLQDELNRNQGRFSSPSQHI 118
      + GL+++ IK + +++ K KE+ K +S E+R++L+DEL++ GRF+ S
Sbjct: 62  LLDQGLNMAGIKQVLLMKENQTEAVKVEETKEISKTELKILRDELQHTGRFNRTSLRQ 121

45  Query: 119 GNM 121
      G++
Sbjct: 122 GDI 124

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4897> which encodes the amino acid sequence <SEQ ID 4898>. Analysis of this protein sequence reveals the following:

```

50      Possible site: 20
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-1770-

The protein has homology with the following sequences in the databases:

```
>GP:BAA00402 GB:D00513 ORF129 [Bacillus cereus]
Identities = 59/122 (48%), Positives = 83/122 (67%), Gaps = 5/122 (4%)

5   Query: 4   KELRRSMAVFPITVMTLTDL SARQIRYYEDQGLIKPERTQGNRRMFSLNDMDRLLEIKD 63
      KE RRS +FPIG VM LT LSARQIRYYE+ L+ P RT+GNRR+FS ND+D+LLEIKD
      Sbjct: 2   KEDRRSAPLFPIGIVMDLTQL SARQIRYYEEHNLVSPTRTKGNRRLFSFNDVDKLEIKD 61

10  Query: 64   FLSEGLNIAAIAIKREYVERQG-----KLMQKQKALTDADVRRILHDEMLTQSGFSTPSQHI 118
      L +GLN+A IK+ + ++ K+ ++ K ++ ++R+IL DE+ F+ S
      Sbjct: 62   LLDQGLNIMAGIKQVILMKENQTEAVKVKEETKEISKTELKILRDELOHTGRFNRTSLRQ 121

      Query: 119 GN 120
      G+
15  Sbjct: 122 GD 123
```

An alignment of the GAS and GBS proteins is shown below.

Identities = 90/123 (73%), Positives = 108/123 (87%)

```
20  Query: 1   MKERELRRRTMAVFPIGAVMKLTDLTARQIRYYEDQGLITPERTEGNRRMFSLNDMDRLLE 60
      MKE+ELRR+MAVFPIG VM LTDL+ARQIRYYEDQGLI PERT+GNRRMFSLNDMDRLLE
      Sbjct: 1   MKEKELRRSMAVFPITVMTLTDL SARQIRYYEDQGLIKPERTQGNRRMFSLNDMDRLLE 60

25  Query: 61   IKDFISDGLHISDIKNEYMQRQHSKEKQKSLSDAEVRRLLQDELNRNQGRFSSPSQHIGN 120
      IKDF+S+GL+I+ IK EY++RQ K +KQK+L+DA+VRR+L DE+ Q FS+PSQHIGN
      Sbjct: 61   IKDFLSEGLNIAAIAIKREYVERQKLMQKQKALTDADVRRILHDEMLTQSGFSTPSQHIGN 120

      Query: 121 MHL 123
      +
30  Sbjct: 121 FRI 123
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1585

35 A DNA sequence (GBSx1679) was identified in *S.agalactiae* <SEQ ID 4899> which encodes the amino acid sequence <SEQ ID 4900>. This protein is predicted to be glutamine synthetase (glnA). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

40  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2157(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4901> which encodes the amino acid sequence <SEQ ID 4902>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

50  INTEGRAL Likelihood = -0.00 Transmembrane 347 - 363 ( 347 - 363)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
55  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

-1771-

Identities = 392/448 (87%), Positives = 421/448 (93%)

Query: 1 MTITAEDIRREVKEKNVTFRLRMFTDILGVMKNVEIPATDEQLDKVLSNKAMFDGSSIEG 60
 M IT DIRREVKEKNVTFRLRMFTDI+GVMKNVEIPAT EQLDKVLSNK MFDGSSIEG
 Sbjet: 1 MAITVADIRREVKEKNVTFRLRMFTDIMGVMKNVEIPATKEQLDKVLSNKVMFDGSSIEG 60

Query: 61 FVRINESDMYLYPDLDTWIVFPWGDENGAVAGLICDIYTAEGEPFAGDPRGNLKRNMKRM 120
 FVRINESDMYLYPDLDTWIVFPWGDENGAVAGLICDIYTAEG+PFAGDPRGNLKR +K M
 Sbjet: 61 FVRINESDMYLYPDLDTWIVFPWGDENGAVAGLICDIYTAEGKPFAGDPRGNLKRALKHM 120

Query: 121 QEMGYKSFNLGPEPEFFFLFKMDENGNTPLDVNDKGGYFDLAPTDLADNTRREIVNVLTQM 180
 E+GYKSFNLGPEPEFFFLFKMD+ GNPTL+VND GGYFDLAP DLADNTRREIVN+LT+M
 Sbjet: 121 NEIGYKSFNLGPEPEFFFLFKMDDKGNPTLEVNDNGGYFDLAPIDLADNTRREIVNILTQM 180

Query: 181 GFEVEASHHEVAVGQHEIDFKYDDVLKACDNIQLFKLVVKTILARKHGLYATFMAKPKFGI 240
 GFEVEASHHEVAVGQHEIDFKY DVLKACDNIQ+FKLVVKTILAR+HGLYATFMAKPKFGI
 Sbjet: 181 GFEVEASHHEVAVGQHEIDFKYADVLKACDNIQIFKLVVKTILAREHGLYATFMAKPKFGI 240

Query: 241 NGSGMHCNMSLFDNNEGNNAFDPEDPRGMQLSEDAYYFLGGLMKHAYNYTAIINPTVNSY 300
 GSGMHCNMSLFDN+GNNAF+D D RGMQLSEDAYYFLGGLMKHAYNYTAI NPTVNSY
 Sbjet: 241 AGSGMHCNMSLFDNQGNNAFYDEADKRGMLSEDAYYFLGGLMKHAYNYTAITNPTVNSY 300

Query: 301 KRLVPGYEAPVYVAVAGNRNSPLIRVPASRGMGTRLELRSDPTANPYLALSVLGGGLE 360
 KRLVPGYEAPVYVAVAG NRSPILRVPASRGMGTRLELRSDPTANPYLAL+VLL +GL+
 Sbjet: 301 KRLVPGYEAPVYVAVAGNSRSPILRVPASRGMGTRLELRSDPTANPYLALAVILLEAGLD 360

Query: 361 GIENKIEAPEPIETNIYAMTVEERRQAGIVDLPSTLHNALEALEEDEVVKAALGTHIYTN 420
 GI NKIEAPEP+E NIY MT+EER +AGI+DLPSLHNAL+AL++D+VV+ ALG HIYTN
 Sbjet: 361 GIINKIEAPEPVEANIYTMTEERNEAGIIDLPSTLHNALEALEEDEVVKAALGTHIYTN 420

Query: 421 FLDAKRIEWASYATYVSQWEIDNYLDLY 448
 FL+AKRIEW+SYAT+VSQWEID+Y+ Y
 Sbjet: 421 FLEAKRIEWSSYATYVSQWEIDHYIHNH 448

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1586

A DNA sequence (GBSx1680) was identified in *S.agalactiae* <SEQ ID 4903> which encodes the amino acid sequence <SEQ ID 4904>. This protein is predicted to be SceB precursor. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA66624 GB:X97985 ORF1 [Staphylococcus aureus]
 Identities = 44/119 (36%), Positives = 66/119 (54%), Gaps = 4/119 (3%)

Query: 26 SFASTNADANTYNYAVDVDYLASAEELIAQHPA-SNTFPLGQCTWGVKE-MATWAGNWWG 83
 S AS + +N + ++ I+ + + SN + GQCT+ V + + G+ WG
 Sbjet: 117 SGASYSTTSNNVHVITTAAPSSNGRSISNGYASGSNLYTSGQCTYYYVFDVGGKIGSTWG 176

Query: 84 NGGDWAASAASADYTVGTQPRVGSIVCWTGDSYGHVAYVTAVDPTNKIQVLESNYAGH 142
 N +WA +AAS+ YTV P+VG+I+ T G YGHVAYV V+ ++V E NY GH
 Sbjet: 177 NASNWANAAASSGYTVNNTPKVGAIMQTQGYGHVAYVEGVNS-NGSVRVSEMNH-GH 233

-1772-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1013> which encodes the amino acid sequence <SEQ ID 1014>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 60/115 (52%), Positives = 81/115 (70%), Gaps = 7/115 (6%)

Query: 55 AHPASNTFPLGQCTWGVKEMATWAGNWWGNGGDWAASAASADYTVGTQPRVGSIVCWTGD 114
++ +SNT+P+GQCTWG K +A WAGN WGNGG WA SA +A Y G+ P VG+I W DG
Sbjct: 291 SYDSSNTYPVGQCTWGA KSLAPWAGNNWGNGGQWASQAAGYRTGSTPMVGAIAVWNDG 350

Query: 115 SYGHVAVYTAVDPTNKIQVLESNYAGHQWIDNYRGWFDPQNTVTPGVVSYIYPN 169
YGHVA V V ++ I+V+ESNY+G Q+I ++RGWF+P V++IYP+
Sbjct: 351 GYGHVAVVVEVQSASS-IRVMESNYSGRQYIADHRGWFNPTG-----VTFIYPH 398

A related GBS gene <SEQ ID 8859> and protein <SEQ ID 8860> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8

McG: Discrim Score: 5.85

GvH: Signal Score (-7.5): 3.11

Possible site: 24

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 0 value: 6.74 threshold: 0.0

PERIPHERAL Likelihood = 6.74 115

modified ALOM score: -1.85

*** Reasoning Step: 3

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

37.5/56.7% over 200aa

Staphylococcus aureus

GP|1340128| ORF1 Insert characterized

ORF00255(376 - 726 of 1107)

GP|1340128|emb|CAA66624.1||X97985(33 - 233 of 255) ORF1 {Staphylococcus aureus}

%Match = 9.0

%Identity = 37.5 %Similarity = 56.7

Matches = 45 Mismatches = 47 Conservative Sub.s = 23

294 324 354 384 414
SVIWI**TRSHQMEENMNKIQKSKTMLGTVALVSASFASFADANTYNYAVDVD-----
: : | : : | : : | : :
MKKIVTATITATAGLATIAFAGHDAQAAEQNNGYNSNDAQSYSTYTTIDAQGNHYHTWTGNWNPSQLTQNN-----
10 20 30 40 50 60 70

462 489 516 546 576 606
-----YLASAEETIAQAHPA-SNTFPLGQCTWGV-KEMATWAGNWWGNGGDWAASAASADYTVGTQ
: : | : : | : : | : : | : : | : : | : : | : :
GSGASYSTTSNNVHVTTPAAPSSNGRSISNGYASGSNLYTSGQCTYYVFDRVGGKIGSTWGNASNWANAASSGYTVNNT
130 140 150 160 170 180 190

-1773-

```

636      666      696      726      756      786      816      846
PRVGSIVCWTGDSYGHVAVYVTAVDPTNKIQVLESNYAGHQWIDNYRGWFDPTVTPGVVSYIYPN*SIKNSSHRRYKS
|:|:|: | | ||||| | : : | | | | : :
5  PKVGAIMQTTQGYGHVAYVEGVNS-NGSVRVSEMN-NGHAGVVTSTRTISANQAGSYNFIH
      210      220      230      240      250

```

SEQ ID 8860 (GBS30) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 8 (lane 2; MW 19.2kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 2; MW 44.2kDa).

GBS30-GST was purified as shown in Figure 193, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1587

A DNA sequence (GBSx1681) was identified in *S.agalactiae* <SEQ ID 4905> which encodes the amino acid sequence <SEQ ID 4906>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -3.93    Transmembrane    2 - 18 ( 1 - 18)
----- Final Results -----
          bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1588

A DNA sequence (GBSx1682) was identified in *S.agalactiae* <SEQ ID 4907> which encodes the amino acid sequence <SEQ ID 4908>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2160(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB06381 GB:AP001516 unknown conserved protein [Bacillus halodurans]
Identities = 353/550 (64%), Positives = 443/550 (80%)

```

```

Query: 6  LKPEEVGVYAIGGLGEIGKNTYIGIEYQDEIIIVDAGIKFPEDDLLGIDYVIPDYSYIVEN 65
      LK +  VYA+GGLGEIGKNTY +++QDEII++DAGIKFPED+LLGIDYVIPDYSY+V+N
Sbjct: 4  LKNNQTAVYALGGLGEIGKNTYAVQFQDEIILLIDAGIKFPEDELLGIDYVIPDYSYLVKN 63

```

```

Query: 66  IDRIKALVITHGHEDHIGGIPFLKQANLPIYAGPLALALIKGKLEEHGLLRDATLYEIH 125
      ++IK L ITHGHEDHIGGIP+LL++ N+PIY G LAL L++GKLEEHGLLR A L++I
Sbjct: 64  ENKIKGLFITHGHEDHIGGIPYLLREVNIPIYGGKLAIGLLRGKLEEHGLLRKAKLHDIQ 123

```

-1774-

5
 Query: 126 ANTELTFKNLSVTFRTTHSIPGLGIVHTPQKVICGDFKFDFTPVGEPADLHRMAA 185
 + + F SV+FFRTTHSIP+ GIV+ TP G ++ TGDFKFDFTPVGEPAL+MA
 Sbjct: 124 EDDIIKFAKTSVSFFRTTHSIPDSYGIVVKTTPGNIVHTGDFKFDFTPVGEPANLTKMAK 183

10
 Query: 186 LGEDGVLCCLSDSTNAEVPTFTNSEKIVGQSIMKIEGIEGRIIFASFASNIFRLQQAEE 245
 +GE+GVLCLSDSTN+E+P FT SE+ VG+SI I +EGRIIFA+FASNI RLQQA E
 Sbjct: 184 IGEEGVLCCLSDSTNSEIPEFTMSERKVGESIDHIFRRVEGRIIFATFASNIHRLQQAEE 243

15
 Query: 246 AAVKTGRKIAVFGSRMEKAIIVNGIELGYIKVPKGTFFIEPSELKNLHASEVLMCTGSQGE 305
 +AV+ GRK+AVFGSRME AI G ELGYIK PK TFIEP++L L +EV+I+CTGSQGE
 Sbjct: 244 SAVRYGRKIAVFGSRMESAINIGQELGYIKAPKNTFIEPNQLNKLDPNEVMILCTGSQGE 303

20
 Query: 306 SMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHGKINNI 365
 MAAL+R+A GTHRQ+ + PGDTVIFSSSPIPGNT SV+K IN + +AG +VIHG +N+I
 Sbjct: 304 PMAALSRAVAFGTHRQIQIIPGDTVIFSSSPIPGNTLSVSKTINQLYKAGANVIHGSLNDI 363

25
 Query: 366 HTSGHGGQEQKLMRLIKPKYFMPVHGEYRMQKVHAGLAVDTGIPKENIFIMENGDVLA 425
 HTSGHGGQ+EQKLMRLIKPKYFMP+HGEYRM K+H LA D G+P EN FIM+NGDVLA
 Sbjct: 364 HTSGHGGQEQKLMRLIKPKYFMPHGEYRMLKMHTKLAEDCGVPAENCFIMDNGDVLA 423

30
 Query: 426 LTSDSARIAGHFNAQDIYVDGNGIGDIGAAVLRDRHDLSEGDVVLAVATVDFDSKMILAG 485
 L D A IAG + +YVDGNGIGDIG VLRDR LSE+G+V+ V +++ + AG
 Sbjct: 424 LHPDEAGIAGKIPSGSVYVDGNGIGDIGNIVLRDRRILSEGLVVVVVSLNMKEYKVITAG 483

35
 Query: 486 PDILSRGFIYMRESGDLIRESQHILFNALRIALKNKDASIQSVNGAIVNALRPFLYEKTE 545
 PD++SRGF+YMRESGDLI+E+Q +L N ++ ++ K + I + L PFLY++T+
 Sbjct: 484 PDLISRGFVYMRESGDLIQAORLLANHLQEVMERKTINQWSEIKNEITDVLGPFLYDRTK 543

40
 Query: 546 REPIIIPMVL 555
 R+P+I+P+++
 Sbjct: 544 RKPMLPIIM 553

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4909> which encodes the amino acid
 sequence <SEQ ID 4910>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.11 Transmembrane 468 - 484 (468 - 484)

----- Final Results -----

bacterial membrane --- Certainty=0.1044(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BA06381 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 353/550 (64%), Positives = 444/550 (80%)

50
 Query: 6 LKPNEVGVFAIGGLGEIGKNTYQIEYQDEIIIVDAGIKFPEDDLLGIDYVIPDYSYIVDN 65
 LK N+ V+A+GGLGEIGKNTY +++QDEII++DAGIKFPED+LLGIDYVIPDYSY+V N
 Sbjct: 4 LKNNQTAVYALGGLGEIGKNTYAVQFQDEIILIDAGIKFPEDELLGIDYVIPDYSYLVKN 63

55
 Query: 66 LDRVKALVITHGHEDHIGGIPFLKQANIPYIYAGPLALALIRGKLEEHGLWREATVYEIN 125
 +++K L ITHGHEDHIGGIP+LL++ NIPIY G LAL L+RGKLEEHGL R+A +++I
 Sbjct: 64 ENKIKGLFITHGHEDHIGGIPYLLREVNPIYGGKLALGLLRGKLEEHGLLRKAKLHDIQ 123

60
 Query: 126 HNTELTFKNMSVTFKTTTHSIPGVGIVHTPQKIICTGDFKFDFTPVGDPADLQMAA 185
 + + F SV+FF+TTHSIP+ GIV+ TP G I+ TGDFKFDFTPVG+PA+L +MA
 Sbjct: 124 EDDIIKFAKTSVSFFRTTHSIPDSYGIVVKTTPGNIVHTGDFKFDFTPVGEPANLTKMAK 183

65
 Query: 186 LGEEGVLCCLSDSTNAEIPFTNSEKVVQSILKIEGIEGRIIFASFASNIYRLQQAEE 245
 +GEEGVLCCLSDSTN+EIP FT SE+ VG+SI I + GRIIFA+FASNI+RLQQA E
 Sbjct: 184 IGEEGVLCCLSDSTNSEIPEFTMSERKVGESIDHIFRRVEGRIIFATFASNIHRLQQAEE 243

-1775-

Query: 246 AAVKTGRKIAVFGRSMEKAIVNGIELGYIKVPGKTFIEPSELKNLHASEVLIMCTGSQGE 305
 +AV+ GRK+AVFGRSME AI G ELGYIK PK TFIEP++L L +EV+I+CTGSQGE
 Sbjct: 244 SAVRYGRKVAVFGRSMESAINIQELGYIKAPKNTFIEPNQLNKLDPNEVMILCTGSQGE 303

5 Query: 306 SMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHGKVNNI 365
 MAAL+R+A GTHRQ+ + PGDTVIFSSSPIPGNT SV+K IN + +AG +VIHG +N+I
 Sbjct: 304 PMAALSRVAFGTHRQIQIIPGDTVIFSSSPIPGNTLSVSKTINQLYKAGANVIHGSLNDI 363

10 Query: 366 HTSGHGGQEQKLMRLSLIKPKYFMPVHGEYRMQKVHAGLAMDIGIPKENIFIMENGVDVLA 425
 HTSGHGGQ+EQKLM LLIKPKYFMP+HGEYRM K+H LA D G+P EN FIM+NGDVLA
 Sbjct: 364 HTSGHGGQEQKLMRLSLIKPKYFMPVHGEYRMLKMHTKLAEDCGVPAENCFIMDNGDVLA 423

15 Query: 426 LTSDSARIAGHFNAQDIYVDGNGIGDIGAAVLRDRRDLSEDGVVLAVATVDFNTQMILAG 485
 L D A IAG + +YVDGNGIGDIG VLRDRR LSE+G+V+ V +++ + AG
 Sbjct: 424 LHPDEAGIAGKIPSGSVYVDGNGIGDIGNIVLRDRRILSEEGLVVVVVSLNMKEYKVTAG 483

20 Query: 486 PDILSRGFIYMRESGDLIRESQRVLFNAIRIALKNKDASIQSVNGAIVNALRPFLYEKTE 545
 PD++SRGF+YMRESGDLI+E+QR+L N ++ ++ K + I + L PFLY++T+
 Sbjct: 484 PDLISRGFVYMRESGDLIQEAQRLLANHLQEVMERKTNQWSEIKNEITDVLGPFLYDRTK 543

Query: 546 REPIIIPMVL 555
 R+P+I+P+++
 Sbjct: 544 RKPMILPIIM 553

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 523/559 (93%), Positives = 550/559 (97%)

Query: 1 MSNINLKPPEEVGVYAIGGLGEIGKNTYGIEYQDEIIIVDAGIKFPEDDLLGIDYVIPDYS 60
 M+NI+LKP EVGV+AIIGGLGEIGKNTYGIEYQDEIIIVDAGIKFPEDDLLGIDYVIPDYS
 30 Sbjct: 1 MTNISLKPNEVGVAIGGLGEIGKNTYGIEYQDEIIIVDAGIKFPEDDLLGIDYVIPDYS 60

Query: 61 YIVENIDRIKALVITHGHEDHIGGIPFLKQANLPIYAGPLALALIKGLEEHGLLRDAT 120
 YIV+N+DR+KALVITHGHEDHIGGIPFLKQAN+PIYAGPLALALI+KGLEEHGL R+AT
 35 Sbjct: 61 YIVDNLDRVKALVITHGHEDHIGGIPFLKQANIPYAGPLALALIRKLEEHGLWREAT 120

Query: 121 LYEIHANTELTFKNLSVTFFRTHHSIPEPLGIVIHPTQGKVICTGDFKFDFTPVGEPADL 180
 +YEI+ NTELTFFKN+SVTFF+TTHSIPEP+GIVIHPTQGK+ICTGDFKFDFTPVG+PADL
 40 Sbjct: 121 VYEINHNTELTFKNMSVTFFKTHHSIPEPVGIVIHPTQGKIICTGDFKFDFTPVGDPADL 180

Query: 181 HRMAALGEDGVLCLLSDSTNAEVPFTFNSEKIVGQSIMKIEGIEGRIIFASFASNIFRL 240
 RMAALGE+GVLCLLSDSTNAE+PTFTNSEK+VGQSI+KIEGI GRIIFASFASNI+RL
 45 Sbjct: 181 QRMAALGEEGVLCLLSDSTNAEIPFTFNSEKVVGQSILKIEGIEHGRIFASFASNIYRL 240

Query: 241 QQAEEAAVKTGRKIAVFGRSMEKAIVNGIELGYIKVPGKTFIEPSELKNLHASEVLIMCT 300
 QQAEEAAVKTGRKIAVFGRSMEKAIVNGIELGYIKVPGKTFIEPSELKNLHASEVLIMCT
 50 Sbjct: 241 QQAEEAAVKTGRKIAVFGRSMEKAIVNGIELGYIKVPGKTFIEPSELKNLHASEVLIMCT 300

Query: 301 GSQGESMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHG 360
 GSQGESMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHG
 55 Sbjct: 301 GSQGESMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHG 360

Query: 361 KNNIHTSGHGGQEQKLMRLSLIKPKYFMPVHGEYRMQKVHAGLAVDTGIPKENIFIMEN 420
 K+NNIHTSGHGGQEQKLM LLIKPKYFMPVHGEYRMQKVHAGLA+D GIPKENIFIMEN
 60 Sbjct: 361 KNNIHTSGHGGQEQKLMRLSLIKPKYFMPVHGEYRMQKVHAGLAMDIGIPKENIFIMEN 420

Query: 421 GDVLALTSDSARIAGHFNAQDIYVDGNGIGDIGAAVLRDRHDLSEDGVVLAVATVDFDSK 480
 GDVLALTSDSARIAGHFNAQDIYVDGNGIGDIGAAVLRDR DLSEDGVVLAVATVDF+++
 65 Sbjct: 421 GDVLALTSDSARIAGHFNAQDIYVDGNGIGDIGAAVLRDRRDLSEDGVVLAVATVDFNTQ 480

Query: 481 MILAGPDILSRGFIYMRESGDLIRESQHILFNAIRIALKNKDASIQSVNGAIVNALRPFL 540
 MILAGPDILSRGFIYMRESGDLIRESQ +LFNAIRIALKNKDASIQSVNGAIVNALRPFL
 Sbjct: 481 MILAGPDILSRGFIYMRESGDLIRESQRVLFNAIRIALKNKDASIQSVNGAIVNALRPFL 540

Query: 541 YEKTEREPIIIPMVLTPDK 559
 YEKTEREPIIIPMVLTPDK
 Sbjct: 541 YEKTEREPIIIPMVLTPDK 559

-1776-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1589

A DNA sequence (GBSx1683) was identified in *S.agalactiae* <SEQ ID 4911> which encodes the amino acid sequence <SEQ ID 4912>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2932(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAB13327 GB:Z99111 ykzG [Bacillus subtilis]
 Identities = 27/75 (36%), Positives = 44/75 (58%), Gaps = 7/75 (9%)

 Query: 1 MIYKVFYQETKERNPRREQTKTLTYVTIDAANELEGRIAARKLVEENTAYNIEFIELLSDK 60
 MIYKVFYQE + P RE+T +LY+ + ++ ++ +K +NIEFI +
20 Sbjct: 1 MIYKVFYQEKADVPVREKTDLSLYIEGVSERDVRTKLKEK-----FNIEFITPVDGA 53

 Query: 61 HLEYEKETGVFELTE 75
 LEYE+++ F++ E
25 Sbjct: 54 FLEYEQQSENFKVLE 68

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4913> which encodes the amino acid sequence <SEQ ID 4914>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3428(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 60/76 (78%), Positives = 70/76 (91%)

40 Query: 1 MIYKVFYQETKERNPRREQTKTLTYVTIDAANELEGRIAARKLVEENTAYNIEFIELLSDK 60
 MIYKVFYQETK+++PRRE TK LY+ IDA +EL+GRI AR+LVE+NT YN+EFIELLSDK
 Sbjct: 1 MIYKVFYQETKDQSPRRESTKALYLNIDATDELDGRIKARRLVEDNTYYNVEFIELLSDK 60

 Query: 61 HLEYEKETGVFELTEF 76
 HL+YEKETGVFELTEF
45 Sbjct: 61 HLDYEKETGVFELTEF 76

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1590

50 A DNA sequence (GBSx1684) was identified in *S.agalactiae* <SEQ ID 4915> which encodes the amino acid sequence <SEQ ID 4916>. This protein is predicted to be glycoprotein endopeptidase. Analysis of this protein sequence reveals the following:

Possible site: 13

-1777-

>>> Seems to have no N-terminal signal sequence (or aa 1-17)

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.0430(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAA76861 GB:Y17797 hypothetical protein [Enterococcus faecalis]
 Identities = 94/182 (51%), Positives = 127/182 (69%), Gaps = 6/182 (3%)

Query: 2 MKVLAFDTSSKALSVAVLNNMECLATVTINIKKNHSINLMPAIDFLMQSIDLEPQDLDR 61
 +++LA DTS++ LS+AV N + L + T +K+NHS+ LMPAID+LM ++L P +DR
 Sbjct: 13 VRILAI DTSNQTLSI AVCENQKILG SYTATV KRNHSLT LMPAIDY LMSQLNLAPT AIDRF 72

15 Query: 62 VVAEGPGSYTGLRVAVATAKMLAYTLKIDLVGVSSLYAL-TNGFSENDLLVPLIDARRNN 120
 VVAEGPGSYTGLR+ V TAK LAYTLK +LVG+SSL AL N + L+VPL DARR N
 Sbjct: 73 VVAEGPGSYTGLRLGVTTAKT LAYTLK KKLVLG ISSLQALANCVGQTGLIVPLFDARRKN 132

20 Query: 121 VYVGIFYQNGDTV----KPDCHTSLEEVLQEVGNKANVHFVGE-VAAFFDQIKKALPHAKI 175
 VY G Y+ D V PD H SL E+L+++ N+ N+ FVGE V F ++I + +PH +I
 Sbjct: 133 VYAGAYRFVDGVWQNELPDQHISLRELLEQLKNEPNLFFVGEDVEKFTEEIAQIIPHGEI 192

25 Query: 176 TE 177
 +
 Sbjct: 193 CD 194

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4917> which encodes the amino acid sequence <SEQ ID 4918>. Analysis of this protein sequence reveals the following:

30 Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.38 Transmembrane 99 - 115 (99 - 115)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9159> which encodes the amino acid sequence <SEQ ID 9160>. Analysis of this protein sequence reveals the following:

40 Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.38 Transmembrane 88 - 104 (88 - 104)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below.

 Identities = 134/232 (57%), Positives = 172/232 (73%), Gaps = 3/232 (1%)

Query: 2 MKVLAFDTSSKALSVAVLNNMECLATVTINIKKNHSINLMPAIDFLMQSIDLEPQDLDR 61
 MK LAFDTS+K LS+A+L++ LA +T+NI+K HS++LMPAIDFLM DL+PQDL+RI
 55 Sbjct: 12 MKTLAFDTSNKTLSLA ILDD ETL LADMT LNIQKKHSVSLMPAIDFLMTCTDLKPQDLERI 71

Query: 62 VVAEGPGSYTGLRVAVATAKMLAYTLKIDLVGVSSLYAL-TNGFSE--NDLLVPLIDARR 118
 VVA+GPGSYTGLRVAVATAK LAY+L I LVG+SSLYAL + N L+VPLIDARR
 Sbjct: 72 VVAKGPGSYTGLRVAVATAKT LAYSLNIALVGISSLYALAASTCKQYPNTLVVPLIDARR 131

60 Query: 119 NNVYVGIFYQNGDTV KPDCHTSLEEVLQEVGNKANVHFVGEVAAFFDQIKKALPHAKITET 178
 N YVG+Y+ G +V P H SLE +++++ + + FVGE A F ++I+K LP A + T

-1778-

Sbjct: 132 QNAYVGYRQGKSVMPQAHASLEVIIEQLVEEGQLIFVGETAPFAEKIQKKLPQAILLPT 191

Query: 179 LPCAVAIGRKQGKMKSVNVDVFPVRYLKRVEAEENWLKNHCETNTEEYIKRV 230

LP A G GQ + NVDAFVP+YLKRVEAEENWLK++ + Y+KR+

Sbjct: 192 LPSAYECGLLGQSLAPENVDAFVPQYLRVEAEENWLKDNELKDDSHYVKRI 243

SEQ ID 4916 (GBS69) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 9; MW 28.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 20 (lane 4; MW 53.9kDa).

The GBS69-GST fusion product was purified (Figure 197, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 285), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1591

A DNA sequence (GBSx1685) was identified in *S.agalactiae* <SEQ ID 4919> which encodes the amino acid sequence <SEQ ID 4920>. This protein is predicted to be ribosomal-protein-alanine acetyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10137> which encodes amino acid sequence <SEQ ID 10138> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC06803 GB:AE000696 ribosomal-protein-alanine acetyltransferase

[Aquifex aeolicus]

Identities = 44/141 (31%), Positives = 74/141 (52%), Gaps = 8/141 (5%)

Query: 9 LREFEMESSEQALAIWSVLSVDYDKSPWSLSQISEDLLKDDSTDYFFVYNDGEVIGFLALQ 68

+RE E E E+ ++ + + + WS +D + + F + DG+V+G++

Sbjct: 4 VREMEREDVER---VYEINRESFTTDAWSRFSFEKDFENKFSRRFVLEEDGKVVGYYVIFW 60

Query: 69 QLVGEVEITNIAVKKNYQGGYAYQLM----SMIADIEVPVFLEVRYSNIVAQKLYERCG 124

+ E I A+ Y+GKGY +L+ S + D V L+VR SN+ A LY++ G

Sbjct: 61 VVKEEATIMTFALAPGYRGKGYGEKLLREAI SRLGDKVKRVVLDVRKSNLRINLYKKLG 120

Query: 125 FVVLRRKKNYYHDPIDAI VM 145

F V+ +RK YY D E+A++M

Sbjct: 121 FKVVTERRKGYSDG-ENALLM 140

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4921> which encodes the amino acid sequence <SEQ ID 4922>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3800(Affirmative) < succ>

-1779-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 65/140 (46%), Positives = 96/140 (68%), Gaps = 1/140 (0%)
 Query: 9 LREFEMES-SEQALAIWSVLSDVYDKSPWSLSQISEDLKKDSTDYFFVYNDGEVIGFLAL 67
 L E M++ EQA I+ +L VY SPW+L Q+ D+++D TDYF +Y+ +++GFLA+
 10 Sbjct: 6 LSESNMKTIVERQAKNIYQLLEMVYGTSPWTLQVLLIDIRRDQTDYFLLYDHDKLLGFLAI 65
 Query: 68 QQLVGEVEITNIAVKKNYQKGAYQLMSMIADIEVPVFLEVRYSNIVAQKLYERC GFVV 127
 Q L GEVE+T IA+ ++Q G A QLM+ + IE +FLEVR SN AQ LY++ GF
 Sbjct: 66 QDLAGEVEMTQIAILPSHQELGLASQLMTHLDSIESDIFLEVRESNHRAQGLYQKFGFKF 125
 15 Query: 128 LKRKKNYYHDPIDAIVMRK 147
 + KR +YY +PIE A++M++
 Sbjct: 126 IGRPDYRNPJETALLMKR 145

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 20 vaccines or diagnostics.

Example 1592

A DNA sequence (GBSx1686) was identified in *S.agalactiae* <SEQ ID 4923> which encodes the amino acid sequence <SEQ ID 4924>. Analysis of this protein sequence reveals the following:

Possible site: 21
 25 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0334 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 35 vaccines or diagnostics.

Example 1593

A DNA sequence (GBSx1687) was identified in *S.agalactiae* <SEQ ID 4925> which encodes the amino acid sequence <SEQ ID 4926>. Analysis of this protein sequence reveals the following:

Possible site: 38
 40 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.75 Transmembrane 86 - 102 (86 - 104)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1702 (Affirmative) < succ>
 45 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04267 GB:AP001508 glycoprotein endopeptidase [Bacillus halodurans]
 50 Identities = 194/331 (58%), Positives = 263/331 (78%), Gaps = 1/331 (0%)
 Query: 6 ILAVESSCDETSVAILKNDKELLANIIASQVESHKRFGGVVPEVASRHHVEVVTTCFEDA 65
 ILA+E+SCDETS A++++N +L+N+++SQ++SHKRFGGVVPE+ASRHHVE +T E+A

-1780-

Sbjct: 12 ILAIETSCDETSAAVIENGTTILSNVSSQIDSHKRFGGVPEIASRHHVEQITVIVEEA 71

Query: 66 LQEAGIVASDLDAAVAVTYGPGLVGALLVGMMAAKAFAWANKLPLIPINHMAGHLMAARDV 125
+ EAG+ +DL AVAVT GPGLVGALL+G+ AAKA A+A++LPLI ++H+AGH+ A R +

5 Sbjct: 72 MHEAGVDFADLAAVAVTEGPGLVGALLIGVNAAKAIAFAHQPLIGVHHIAGHIYANRLL 131

Query: 126 KELQYPLIALLVSGGHTELVYVSEPGDYKIVGETRDDAVGEAYDKVGRVMGLTYPAGREI 185
KEL++PLIAL+VSGGHTEL+Y+ G++++GETRDDAVGEAYDKV R +GL YP G I

10 Sbjct: 132 KELEFPLLALVSGGHTELIYMENHGEFEVIGETRDDAVGEAYDKVARTLGLPYPGGPHI 191

Query: 186 DQLAHKGQDTYHFPRA MIKEDHLEFSFSGLKSAFINLHHNAEQKGEALVLEDLCASFQAA 245
D+LA G+DT FPRA ++ D +FSFSGLKSA IN HNA+Q+GE + ED+ ASFQA+

Sbjct: 192 DRLAVNGEDTLQFPRAWLEPDSFDFSGLKSAVINTLHNAKQRGENVQAEDVAASFQAS 251

15 Query: 246 VLDILLAKTQKALKYPVKTLVAGGVAANOGLRERLATDISPD-IDVVIPLRLCGDNA 304
V+D+L+ KT+KA +Y V+ +++AGGVAAN+GLR L + ID+VIPPL LC DNA

Sbjct: 252 VIDVLVTTKKAAEYKVRQVLLAGGVAANKGLRTALEEAFFKEPIDLVIPPLSLCTDNA 311

Query: 305 GMIALAAAEFEKENFASLKLNKPSLAFES 335
MI AA+I+F+++ FA + LN +PSL E+

20 Sbjct: 312 AMIGAAASIKFKQQTFA GMDLNGQPSLELEN 342

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4927> which encodes the amino acid sequence <SEQ ID 4928>. Analysis of this protein sequence reveals the following:

25 Possible site: 38
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -2.76 Transmembrane 86 - 102 (85 - 104)

30 ----- Final Results -----
bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35 >GP:BAB04267 GB:AP001508 glycoprotein endopeptidase [*Bacillus halodurans*]
Identities = 196/330 (59%), Positives = 255/330 (76%), Gaps = 2/330 (0%)

Query: 6 ILAVESSCDETSVAILKNESTLLSNVIASQVESHKRFGGVPEVASRHHVEVITTCFEDA 65
ILA+E+SCDETS A+++N +T+LSNV++SQ++SHKRFGGVPE+ASRHHVE IT E+A

40 Sbjct: 12 ILAIETSCDETSAAVIENGTTILSNVSSQIDSHKRFGGVPEIASRHHVEQITVIVEEA 71

Query: 66 LQEAGISASDLSAVAVTYGPGLVGALLVGLAAAKAFAWANHLPLIPVNHMAGHLMAAREQ 125
+ EAG+ +DL+AVAVT GPGLVGALL+G+ AAKA A+A+ LPLI V+H+AGH+ A R

45 Sbjct: 72 MHEAGVDFADLAAVAVTEGPGLVGALLIGVNAAKAIAFAHQPLIGVHHIAGHIYANRLL 131

Query: 126 KPLVYPLIALLVSGGHTELVYVPEPGDYHIIGETRDDAVGEAYDKVGRVMGLTYPAGREI 185
K L +PL+AL+VSGGHTEL+Y+ G++ +IGETRDDAVGEAYDKV R +GL YP G I

Sbjct: 132 KELEFPLLALVSGGHTELIYMENHGEFEVIGETRDDAVGEAYDKVARTLGLPYPGGPHI 191

50 Query: 186 DQLAHKGQDTYHFPRA MITEDHLEFSFSGLKSAFINLHHNAKQKGELELLEDLCASFQAA 245
D+LA G+DT FPRA + D +FSFSGLKSA IN HNAKQ+G+ + ED+ ASFQA+

Sbjct: 192 DRLAVNGEDTLQFPRAWLEPDSFDFSGLKSAVINTLHNAKQRGENVQAEDVAASFQAS 251

55 Query: 246 VLDILLAKTKKALSRYPAKMLVAGGVAANOGLRDLAQEI--THIEVVIPKLRLCGDNA 303
V+D+L+ KTKKA Y + +++AGGVAAN+GLR L + I++VIP L LC DNA

Sbjct: 252 VIDVLVTTKKAAEYKVRQVLLAGGVAANKGLRTALEEAFFKEPIDLVIPPLSLCTDNA 311

Query: 304 GMIALAAAEYDKQHFANMSLNKPSLAFD 333
MI AA+I++ +Q FA M LN +PSL +

60 Sbjct: 312 AMIGAAASIKFKQQTFA GMDLNGQPSLELE 341

An alignment of the GAS and GBS proteins is shown below.

Identities = 288/334 (86%), Positives = 313/334 (93%), Gaps = 1/334 (0%)

-1781-

Query: 1 MIDRYILAVESSCDETSVAILKNDKELLANIIASQVESHKRFGGVVPEVASRHHVEVVT 60
 M DRYILAVESSCDETSVAILKN+ LL+N+IASQVESHKRFGGVVPEVASRHHVEV+TT
 Sbjct: 1 MTDRYILAVESSCDETSVAILKNESLTLNVIASQVESHKRFGGVVPEVASRHHVEVIT 60

5 Query: 61 CFEDALQEAGIVASDLDAVAVTYGPGLVGALLVGMAAAKAFAWANKLPLIPINHMAGHLM 120
 CFEDALQEAGI ASDL AVAVTYGPGLVGALLVG+AAAKAFAWAN LPLIP+NHMAGHLM
 Sbjct: 61 CFEDALQEAGISASDL SAVAVTYGPGLVGALLVGLAAAKAFAWANHLPLIPVNHMAGHLM 120

10 Query: 121 AARDVKELQYPLLALLVSGGHTELVYVSEPGDYKIVGETRDDAVGEAYDKVGRVMGLTYP 180
 AAR+ K L YPL+ALLVSGGHTELVYV EPGDY I+GETRDDAVGEAYDKVGRVMGLTYP
 Sbjct: 121 AAREQKPLVYPLIALLLVSGGHTELVYVPEPGDYHIIGETRDDAVGEAYDKVGRVMGLTYP 180

15 Query: 181 AGREIDQLAHKGQDTYHFPRAMIKEDHLEFSFSGLSAFINLHHNAEQKGEALVLEDLCA 240
 AGREIDQLAHKGQDTYHFPRAMI EDHLEFSFSGLSAFINLHHNA+QKG+ L+LEDLCA
 Sbjct: 181 AGREIDQLAHKGQDTYHFPRAMITEDHLEFSFSGLSAFINLHHNAKQKGELELLEDLCA 240

20 Query: 241 SFQAAVLDILLAKTQKALLKYPVKTLVVAGGVAANQGLRERLATDISPDIDVVIPLRLC 300
 SFQAAVLDILLAKT+KAL +YP K LVVAGGVAANQGLR+RLA +I+ I+VVIP LRLC
 Sbjct: 241 SFQAAVLDILLAKTKKALSRYPAKMLVVAGGVAANQGLRDLRAQEIT-HIEVVIPKLRIC 299

Query: 301 GDNAGMIALAAAIEFEKENFASLKLNAKPSLAFE 334
 GDNAGMIALAAAIE++K++FA++ LNAKPSLAF+
 Sbjct: 300 GDNAGMIALAAAIEYDKQHFANMSLNAKPSLAFD 333

25 SEQ ID 4926 (GBS371) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 7; MW 41kDa), in Figure 170 (lane 4 & 5; MW 55kDa) and in Figure 239 (lane 6; MW 55kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 7; MW 65kDa).

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1594

A DNA sequence (GBSx1688) was identified in *S.agalactiae* <SEQ ID 4929> which encodes the amino acid sequence <SEQ ID 4930>. Analysis of this protein sequence reveals the following:

35 Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1027(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1595

A DNA sequence (GBSx1689) was identified in *S.agalactiae* <SEQ ID 4931> which encodes the amino acid sequence <SEQ ID 4932>. Analysis of this protein sequence reveals the following:

50 Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1307(Affirmative) < succ>

-1782-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1596

10 A DNA sequence (GBSx1690) was identified in *S.agalactiae* <SEQ ID 4933> which encodes the amino acid sequence <SEQ ID 4934>. This protein is predicted to be L4171-60 protein. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.

15 ----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 A related GBS nucleic acid sequence <SEQ ID 10135> which encodes amino acid sequence <SEQ ID 10136> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC24656 GB:AE001274 L4171.5 [Leishmania major]
 Identities = 118/282 (41%), Positives = 167/282 (58%), Gaps = 4/282 (1%)

25 Query: 2 GGTQTNQVVISSMLASYEGVIAAETGHVSSHEAGAIEFSGHKVLTLPSHNGKLLASEVAT 61
 GGTQTN + S L +E VIA + GH+S+HE GAIE +GHKV+T P +GKL ++
 Sbjct: 74 GGTQTNLIACSLALRPWEAVIATQLGHISTHETGAIEATGHKVVITAPCPDGKLRVAD--- 130

30 Query: 62 YIETFYADGNVQHMVFPGMVYISHPTGYGLYSKAELEELSCKICKHYQIPLFIDGARLG 121
 IE+ + +HMV P +VYIS+ TE GT Y+K ELE++S CK + + LF+DGARL
 Sbjct: 131 -IESALHENRSEHMVIPKLVIYISNTTEVGTQYTKQELEDISASCKEHLGLYFLDGLARLAS 189

35 Query: 122 GLAAKDTDVDFPTIAALSDFYIGGKMGALAGEAVVFTKKNRPKQFTTIVKQHGALLAK 181
 L++ D+ IA L+D+FYIG TK G + GEA++ ++KQ GAL+AK
 Sbjct: 190 ALSSPVNDLTLADIARLTDMFYIGATKAGGMFGEALIIINDALKPNARHLIKQRGALMAK 249

40 Query: 182 GRLLGLAFDRFFTDNLYLKIGKHAIDLAEELKIILEEKGYFYLKSPNTNQFFIIVENTKL 241
 G LLG+ F+ DNL+ ++G H+ +A LK LE G S +NQ F I+ENT +
 Sbjct: 250 GWLLGIQFEVLMKDNLFELGAHSNKMAAILKAGLEACGIRLAWPSASNQLFPILNTMI 309

45 Query: 242 ADLAKNVAYSFWKEYDDHHTVIRLATSWSTSREDVTALRNVL 283
 A+L + E D ++RL TSW+T ++ VL
 Sbjct: 310 AELNNDPDMYTVPEPLKDGTCIMRLCTSWATEEKECHRFVEVL 351

No corresponding DNA sequence was identified in *S.pyogenes*.

50 SEQ ID 4934 (GBS648) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 8-10; MW 60kDa) and in Figure 186 (lane 6; MW 60kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 12; MW 35kDa), in Figure 140 (lane 10; MW 35kDa) and in Figure 178 (lane 7; MW 35kDa).

Purified GBS648-GST is shown in Figure 243, lane 6; purified GBS648-His is shown in Fig. 229, lane 7.

-1783-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1597

A DNA sequence (GBSx1691) was identified in *S.agalactiae* <SEQ ID 4935> which encodes the amino acid sequence <SEQ ID 4936>. Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2279(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1598

A DNA sequence (GBSx1692) was identified in *S.agalactiae* <SEQ ID 4937> which encodes the amino acid sequence <SEQ ID 4938>. This protein is predicted to be ribosomal protein S14 (rpsN). Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.3848(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12716 GB:Z99108 similar to ribosomal protein S14 [Bacillus subtilis]
Identities = 67/89 (75%), Positives = 76/89 (85%)

35 Query: 1 MAKKSIAKFQKQQLVEQYAEIRRELKEKGDYEALRKLPKDSNPNRLKNRDLIDGRPHA 60
 MAKKS+AK K+Q+LVEQYA +RRELKEKGDYEAL KLP+DS P RL NR ++ GRP A
 Sbjct: 1 MAKKSVAKEKLRQQLVEQYAGIRRELKEKGDYEALSKLPRDSAPGRLHNRCMVTGRPRA 60

Query: 61 YMRKFGMSRINFRNLAYKGQIPGIKKASW 89

40 YMRKF MSRI FR LA+KGQIPG+KKASW
 Sbjct: 61 YMRKFKMSRIAFRELAHKGQIPGVKKASW 89

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4939> which encodes the amino acid sequence <SEQ ID 4940>. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.3799(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-1784-

Identities = 73/89 (82%), Positives = 85/89 (95%)

Query: 1 MAKKSKIAKFQKQKLVEQYAE LRRELKEKGDYEALRKLPKDSNPNRLKNRDLIDGRPHA 60
 MAKKSKIAK+QKQ +L+EQYA+LRR+LK KGDYE+LRKLP+DSNPNRLKNRD IDGRPHA
 5 Sbjet: 1 MAKKSKIAKYQKQLIEQYADLRDLKAKGDYESLRKLPKDSNPNRLKNRDKIDGRPHA 60

Query: 61 YMRKFGMSRINFRNLAYKGQIPGIKASW 89
 YMRKFG+SRINFR+LA+KGQ+PG+ KASW
 10 Sbjet: 61 YMRKFGVSRINFRDLAHKGQLPGVTKASW 89

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1599

A DNA sequence (GBSx1693) was identified in *S.agalactiae* <SEQ ID 4941> which encodes the amino acid sequence <SEQ ID 4942>. Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5183(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB95931 GB:AL359989 galactose-1-phosphate uridylyltransferase
 [Streptomyces coelicolor A3(2)]
 Identities = 31/105 (29%), Positives = 51/105 (48%), Gaps = 4/105 (3%)

Query: 27 DKCPFC--DKSQLGKILDVKDDMIWVENKYPTL--EETVQTLVIESNDHNGDISVYSESK 82
 D+CP C D +L +I D D++ EN++P+L + +V ++DH+ + SE +
 30 Sbjet: 68 DQCPLCPDGERLSEIPDSAYDVVVFENRFP SLAGDSGRCEVVCFTSDHDASFADLSEEQ 127

Query: 83 MRQLLDYLLSKWQLMEESGHYRSVVLRYNFGPLSGGSLRHPSQI 127
 R +LD + + V + N G G +L HPH QI
 35 Sbjet: 128 ARLVLDATDRTSELSHLPSVEQVFCFENRGAEIGVTLGHPHGQI 172

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1600

A DNA sequence (GBSx1694) was identified in *S.agalactiae* <SEQ ID 4943> which encodes the amino acid sequence <SEQ ID 4944>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10133> which encodes amino acid sequence <SEQ ID 10134> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1785-

>GP: BAB06998 GB: AP001518 unknown conserved protein [Bacillus halodurans]
Identities = 186/410 (45%), Positives = 258/410 (62%), Gaps = 27/410 (6%)

5 Query: 4 YDTIIIGGGPAGMMAAIISSNFYGNKTLTLLIEKNKRLGKKLAGTGGGRCNVTNNGNLDELLA 63
++ I+IGGGPAG+MA++S+ +G + LL++K +LG+KLA +GGGRCNVTN LDEL+A
Sbjct: 2 HEVIVIGGGPAGLMAVSAAEHGARVLLLDKGDKLGRKLAISGGGRCNVTNRMPDLDELIA 61

10 Query: 64 GIPGNRFLYSVFSQFDNHDIIINFFQDNGVTLKEEDHGRMFPTTDSRTIINALENKIKE 123
IPGNRFR+YS FS F+N DII FF+ G+ LKEED GRMFP +DK+ T++ L +I +
Sbjct: 62 HIPGNRFRMYSFVSFVNEDIIIRFFERLGLALKEEDGRMFPVSDKATTVVQITLLKRIND 121

15 Query: 124 LGGQIMTDTEVVSVK-KIGDSFYIKTKDTQFASDK-LIVTTGGKSYPSTGSTGFHDIAR 181
LG + T+T V S++ G ++ K+ + K +IV TGG+S P TGSTG + A+
Sbjct: 122 LGVTVRTNTAVASLEYDDGRIAMVQLKNGERLTKTKTIVATGGQSVPHSTGSTGDAYPWAK 181

20 Query: 233 GLSGPAALRISSFVKGETIY-----LDVLPNISVKEL-EIHFQN---EREKSLKNA 280
GLSGPAALR S +V Y +D+ P I + L + QN E +K+LK
Sbjct: 242 GLSGPAALRCSQYVVKALKKYKQPTIEMRIDLRPTIPAEALFQETIQNIKAEPKALKKTV 301

25 Query: 281 LKILLPERLAEFYAEDL--PEKVKQVSVKD--LEMLIQKLKLPILVTGKMSLAKSFVTK 336
L+ + PER ++ E L + SV+ + ++Q+LK V G +S+ K+FVT
Sbjct: 302 LRGIAPERFLQYIYERLRIDSNLPCASVRHEVIREIVQQLKSFSFHVNGTLSIEKAFVTG 361

30 Query: 337 GGVDLKEINPKTLESKKVAGLHFAGEVLDINAHTGGFNITSALCTGWWAG 386
GGV +KEI PKT+ SKK AGL F GEVLDI+ +TGG+NIT A TG+ AG
Sbjct: 362 GGVSVEKIEPKTMHSSKKKAGLFFCGEVLDIHGYTGGYNITCAFSTGYTAG 411

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4945> which encodes the amino acid sequence <SEQ ID 4946>. Analysis of this protein sequence reveals the following:

35 Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0448(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 308/386 (79%), Positives = 344/386 (88%)

45 Query: 1 MKHYDTIIIGGGPAGMMAAIISSNFYGNKTLTLLIEKNKRLGKKLAGTGGGRCNVTNNGNLDE 60
M YDTIIIGGGPAGMMAAIISS++YG KTLTLLIEKN+RLGKKLAGTGGGRCNVTN+GNLD
Sbjct: 1 MTQYDTIIIGGGPAGMMAAIISSSYGYKTLTLLIEKNRRLGKKLAGTGGGRCNVTNSGNLDV 60

50 Query: 61 LLAGIPGNRFLYSVFSQFDNHDIIINFFQDNGVTLKEEDHGRMFPTTDSRTIINALENK 120
L+AGIPGNRFLYSVFSQFDNHDII FF++NGV LKEEDHGRMFPTTDSRTII+ALE K
Sbjct: 61 LMAGIPGNRFLYSVFSQFDNHDIIAFFEENGVKLKEEDHGRMFPTTDSRTIIDALEKK 120

55 Query: 121 IKELGGQIMTDTEVVSVKKIGDSFYIKTKDTQFASDKLIVTTGGKSYPSTGSTGFHDIAR 180
IK LGQ++T TEVVSVKK D FY+K+ D F KLIVTTGGKSYPSTGSTGFHDIAR
Sbjct: 121 IKALGGQVLTSTEVVSVKKQDDLFLYKLSADQFTTCQKLIVTTGGKSYPSTGSTGFHDIAR 180

60 Query: 181 RHFKEVTDMEAAESPLLTDFFPHKKLQGISLDDVTLSEFKHIIITHDLLFTHFGLSGPAAL 240
RHFKL VID+EAASPLLTDFFPHK LQGISLDDVTL++KH+ITHDLLFTHFGLSGPAAL
Sbjct: 181 RHFKLTVTDLEAAESPLLTDFFPHKVLQGISLDDVTLSDYKHVITHDLLFTHFGLSGPAAL 240

65 Query: 241 RISSFVKGETIYLDVLPNISVKELIHFQNEREKSLKNALKILLPERLAEFYAEDLPEK 300
R+SSFVKGGE LD LP++S +L + ++R+K++KNALK LLPER+A+F +ED PEK
Sbjct: 241 RLSSFVKGGEIAELDFLPHLSTDDL TAYLSDQDKNIKALKGLLPERVADFLSEDYPEK 300

Query: 301 VKQVSVKDLEMLIQKLKLPILVTGKMSLAKSFVTKGGVDLKEINPKTLESKKVAGLHFA 360

-1786-

VKQ+S K + L+ KLK L I +TGKMSLAKSFVTKGGVDLKEINPKTLESKKV GL+FA
 Sbjct: 301 VKQLSPKQEKELLDKLKHLQIPITGKMSLAKSFVTKGGVDLKEINPKTLESKKVPGLYFA 360

Query: 361 GEVLDINAHTGGFNITSALCTGWVAG 386
 GEVLDINAHTGGFNITSALC+GW+AG
 Sbjct: 361 GEVLDINAHTGGFNITSALCSGWIAG 386

SEQ ID 4944 (GBS196) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 3; MW 44.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 4; MW 69.5kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1601

A DNA sequence (GBSx1695) was identified in *S.agalactiae* <SEQ ID 4947> which encodes the amino acid sequence <SEQ ID 4948>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1550(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10131> which encodes amino acid sequence <SEQ ID 10132> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA73267 GB:Y12736 orfX [Lactococcus lactis subsp. cremoris]
 Identities = 51/173 (29%), Positives = 87/173 (49%), Gaps = 20/173 (11%)

Query: 19 KTVSELAEILGVSRQAMNNRV-KTLPEECVEK---NSKGVTVVNRDGLIKLEEIYKKTIL 74
 KT+ ELA+ LGVS+Q + N++ K E+ V+ V+N G + KKT+
 Sbjct: 6 KTIKELADELGVSKQTIRNKIDKDFREKFVQTIKIKGNNTLVINNAGY----SLKKKTILQ 61

Query: 75 EEEPIDEEASRRELLEILVDEKNTETIRLYEQLKAKDIQIASKDEQLHVKDIQIAEKDKQ 134
 + + + + + I L EQL K+ Q++ KD+QL KD QI++
 Sbjct: 62 NDTAQTAQTLQNDTAQTKL-----ICFLEEQLDKKEQQLSVKDKQLENKDTQISQMQL 115

Query: 135 LDQQQQLTLTAMEDTQRLQLELNEAKA-----EVEEIQEAKKEKIQELEAVK 181
 LDQQQ+L L + + + E+NE KA +++++ + E +E+E +K
 Sbjct: 116 LDQQQLALQDKKLLLEEKSEINELKALKMPREDMKDGSSIRGEAQEEIERLK 168

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4949> which encodes the amino acid sequence <SEQ ID 4950>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3951(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 132/194 (68%), Positives = 154/194 (79%), Gaps = 4/194 (2%)

-1787-

Query: 1 MIFFYKCI---STKEEVMTEKTVSELAELGVSRQAMNVRVKTLPPEECVEKNSKGVTVV 57
 M+ F +I S KEE M +EKTVSELA+ILGVSRQA+NNRVK+LPPEE ++KN KGVTVV
 Sbjct: 1 MVLFLIRIFSDSDKEENMGIEKTVSELADILGVSRQAVNNRVKSLPEEDLDKNEKGVTVV 60

5 Query: 58 NRDGLIKLEEIYKKTILEEEPIDEEASRRELLEILVDEKNTETITRLYEQLKAKDIQIASK 117
 R GL+KLEEIYKKTII ++EPI EE +RELLEILVDEKNTETITRLYEQLKAKD Q+ASK
 Sbjct: 61 KRSGLVKLEEIYKKTIFDDEPISEETKQRELEILVDEKNTETITRLYEQLKAKDAQIASK 120

10 Query: 118 DEQLHVKDIQIAEKDKQLDQQQQLTLTAMEDTQRIQLELNEAKAEVEETIQEAKKEKIQEL 177
 DEQ+ VKD+QIAEKDKQLDQQQQLT AM D + L+LEL EAKAE + + + E++Q
 Sbjct: 121 DEQMRVKDVQIAEKDKQLDQQQQLTAKAMADKETLKLELEEAKEANQAR-LQVEEVQAE 179

Query: 178 EAVKKSFFGRFFNK 191
 KK FF R F K
 15 Sbjct: 180 VGPKKGFFTRLFAK 193

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1602

20 A DNA sequence (GBSx1697) was identified in *S.agalactiae* <SEQ ID 4951> which encodes the amino acid sequence <SEQ ID 4952>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2157(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06137 GB:AP001515 DNA polymerase III (alpha subunit)
 [Bacillus halodurans]
 Identities = 31/87 (35%), Positives = 52/87 (59%), Gaps = 1/87 (1%)

35 Query: 13 EYIAPDLEFNTVGE-HSHIIQVSAVKYSNHQEIADFDTYVHTKVPLQSFINGLTGITARD 71
 E++ FD+E + ++ II+++AVK N + I F+ + PL + I LTGIT
 Sbjct: 418 EFVVFVDETGLSAVYNKIIELA AVKVNGEIIDRFERFADPHEPLTNTTIELTGITDDM 477

40 Query: 72 IIGAPKIEIVLTDFQSFVGDTPPLIGYN 98
 + G P++E VL +F +F+GD L+ +N
 Sbjct: 478 LKGQPEVEQVLNEFHAFIGDAVLVAHN 504

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4953> which encodes the amino acid sequence <SEQ ID 4954>. Analysis of this protein sequence reveals the following:

45 Possible site: 29
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3427(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

55 Identities = 136/200 (68%), Positives = 159/200 (79%)

Query: 3 FLGEIMKQLQEYIAPDLEFNTVGEHSHIIQVSAVKYSNHQEIADFDTYVHTKVPLQSFIN 62
 FL E MK L YIAPDLEFNTV + SHIIQVSAVKY +H+E+ FDTYV+T VPLQSFIN
 Sbjct: 9 FLEENMKHLDTYIAPDLEFNTVNDVSHIIQVSAVKYDHHKEVDSFDTYVYTDVPLQSFIN 68

-1788-

Query: 63 GLTGITARDIIGAPKIEIVLTDQSFVGDTPILIGYNGYKSDLPLLVENGLDLTSQYQVDL 122
 GLTGIT+ I PK+E V+ F++FVG+ PLIGYN KSDLP+L ENGLDL QYQ+DL
 Sbjct: 69 GLTGITSKDIAAEPKVEEVMAAFKNFVGELPLIGYNAQKSDLPLAENGLDLRDQYQIDL 128

Query: 123 YDEAFVRRSTDNLNGIVNLKLTTVADFLGIKGAHNSLEDARMTARVYEKFLDLDENKIYL 182
 +DEA+ RRS DLNGI NL+L TVA FLGIK+ HNSLEDARMTA +Y+ FL+ D NK YL
 Sbjct: 129 FDEAYDRRSADLNGIANLRLQTVATFLGIKGRGHNSLEDARMTAVIYKSFLETDINKAYL 188

Query: 183 KQQKEVAVDSPFATLGNLFD 202
 QQ+EV D+PFA LG+ FD
 Sbjct: 189 SQQEEVTTDNPFAALGDFFD 208

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1603

A DNA sequence (GBSx1698) was identified in *S.agalactiae* <SEQ ID 4955> which encodes the amino acid sequence <SEQ ID 4956>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -12.10 Transmembrane 143 - 159 (136 - 166)
 INTEGRAL Likelihood = -4.73 Transmembrane 169 - 185 (168 - 188)

----- Final Results -----

bacterial membrane --- Certainty=0.5840(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB42766 GB:AL049841 transcriptional regulator [Streptomyces
 coelicolor A3(2)]
 Identities = 46/141 (32%), Positives = 71/141 (49%), Gaps = 11/141 (7%)

Query: 5 YSTGDLAKEAGVTVRTVQYYDKRGILSPSELSEGGRVYSIADLEKLRIIYLRDLDFSI 64
 YS G +A AGVTVRT+ +YD G+L PSE S G R YS ADL++L+QI++ R+L F +
 Sbjct: 3 YSVGQVAGFAGVTVRTLHHYDDIGLLVPSESHAGHRRYSADLDRLQQLFYRELGFPL 62

Query: 65 DNIKNLFTEDNASQILELFLQVQIRELRL-----AIDSKDKLDKAVNLLKTVEKQD 116
 D + L + A L Q ++ R+ A++ + +NL ++
 Sbjct: 63 DEVAALLDDPAADPRAHLRRQHELLSARIGKLQKMAAAVEQAMEARSMGINL---TPEEK 119

Query: 117 SKTLGYLSDIVLMEENKRKRWG 137
 + G EE + +WG
 Sbjct: 120 FEVFGDFDPDQYEEVRRERWG 140

There is also homology to SEQ ID 1712.

SEQ ID 4956 (GBS372) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 8; MW 55kDa).

GBS372-GST was purified as shown in Figure 215, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1789-

Example 1604

A DNA sequence (GBSx1699) was identified in *S.agalactiae* <SEQ ID 4957> which encodes the amino acid sequence <SEQ ID 4958>. This protein is predicted to be cyclopropane-fatty-acyl-phospholipid synthase (mma2). Analysis of this protein sequence reveals the following:

```

5   Possible site: 53
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3145(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AAD07482 GB:AE000557 cyclopropane fatty acid synthase (cfa)
    [Helicobacter pylori 26695]
    Identities = 167/397 (42%), Positives = 254/397 (63%), Gaps = 14/397 (3%)

    Query: 2   VMSLIIKQLIKSTFDIPLQVTYPNGNIETYNGSNPHVKLKNKNSVSELSKDPISIVLG 61
              ++  ++K + K + QV + + ++ +P LK+++ S++ KD S+ +
20  Sbjct: 1   MISKFLLKSMFKQWKNGDYQVFWDNSVYRNGEHSKFTLKIHRPLKFSDIKKDMSLTIA 60

    Query: 62   EAVMDGDIEIYGSIQELILSAY-RCGDSFLRNSKFSKLIPKQFHDKKHKSQDIQKHIDIG 120
              EA MDG I+I GS+ E++ S Y + L +K I K + S+I KHYD+G
25  Sbjct: 61   EAYMDGVIDIEGSMDEVMSLSYLQNTYEHLLHKHDNAKAIQKPIKES----SNISKHYDLG 116

    Query: 121  NDFYKLWLDGMTYSCAYFKHENDSLEQAQLNKVHHILNKLNAQPGGKLLDIGCGWGTLI 180
              NDFY +WLD+T++YSCAYFK ++D+L AQL K+ H L KL+ +PG KLLDIGCGWG L
30  Sbjct: 117  NDFYSIWLDETLISYSCAYFKDDDTLHAAQLQKLDHTLKKLHLKPGEKLLDIGCGWGYLS 176

    Query: 181  ITAAKEYGLNATGITLSEEQASFITKRIKEEGLENKVTVLIDYRDI---RETYDYITSV 237
              + AA+EYG GIT+S EQ KR++E GLE+KVT+ + +Y+D+ +D + SV
35  Sbjct: 177  VKAAQEYGAEVMGITISSEQYKQANKRVQELGLEDKVTIKLLNYQDLDGRLYRFDKVVSV 236

    Query: 238  GMFEHVGKENLSQYFQTISKRLNINGLALIHGITGVGGNHGSGTNSWINKYIFPGGYIP 297
              GMFEHVGK+NL YF+ + + L G+ L+H I G TN+W++KYIFPGGY+P
40  Sbjct: 237  GMFEHVGKDNLPFYFKKVEVLKRGGMFLLHSILCCFEGK---TNAWVDKYIFPGGYLP 292

    Query: 298  RLITENLNHIASAGLQIADLEPLRRHYQKTLLELWTKNFHNLPEVQK-THDKRFINMWDLY 356
              L E ++ ++ + E LR HY KTL++W NF++ L +V++ ++D+RFI MWDLY
45  Sbjct: 293  SLREVMSVMSECDPFLHLLMAESLRIHYAKTLDIWRNNFNHNLQVKKLSYDERFIRMWDLY 352

    Query: 357  LQSCAASFESGNIDIFQYLLSKGVSKDTMPMTRDYM 393
              L++CA++F G+ D+FQ LL+ V +T P+T++Y+Y
50  Sbjct: 353  LRTCASAFRVGSADLFQLLLTNSVD-NTFPLTKEYIY 388

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1605

50 A DNA sequence (GBSx1700) was identified in *S.agalactiae* <SEQ ID 4959> which encodes the amino acid sequence <SEQ ID 4960>. Analysis of this protein sequence reveals the following:

```

    Possible site: 35
    >>> Seems to have no N-terminal signal sequence

55  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4903(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-1790-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11796 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
Identities = 44/97 (45%), Positives = 60/97 (61%)

Query: 1 MMNMQNMMRQAQKLQKQMEQKQADLAASQFTGKSAQELVTFTGDKKLISIDYKEAVVD 60
M NMQ MM+Q QK+QK M + Q +LA G + +VTV G K+++ + KE VVD
Sbjct: 5 MGNMQKMMKQMOKMQDMAKAEKVVETAGGGMVTVKANGQKEILDVIIKEEVVD 64

Query: 61 PEDIETLQDMTTQAINDALSQVDDATKKIMGAFAGKM 97
PEDI+ LQD+ A N+AL +VD+ T + MG F M
Sbjct: 65 PEDIDMLQDLVLAATNEALKKVDEITNETMGQFTKGM 101

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4961> which encodes the amino acid sequence <SEQ ID 4962>. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4451(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 84/99 (84%), Positives = 94/99 (94%)

Query: 1 MMNMQNMMRQAQKLQKQMEQKQADLAASQFTGKSAQELVTFTGDKKLISIDYKEAVVD 60
MMNMQNMM+QAQKLQKQMEQKQADLAA QFTGKSAQ+LVT TFTGDKKL+ ID+KEAVVD
Sbjct: 1 MMNMQNMMKQAQKLQKQMEQKQADLAAMQFTGKSAQDLVTATFTGDKKLVGIDFKEAVVD 60

Query: 61 PEDIETLQDMTTQAINDALSQVDDATKKIMGAFAGKMPF 99
PED+ETLQDMTTQAINDAL+Q+D+ TTK +GAFAGK+PF
Sbjct: 61 PEDVETLQDMTTQAINDALTQIDETTKKTLGAFAGKLPF 99

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1606

A DNA sequence (GBSx1701) was identified in *S.agalactiae* <SEQ ID 4963> which encodes the amino acid sequence <SEQ ID 4964>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3963(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1791-

Example 1607

A DNA sequence (GBSx1702) was identified in *S.agalactiae* <SEQ ID 4965> which encodes the amino acid sequence <SEQ ID 4966>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 48
      >>> Seems to have no N-terminal signal sequence
          INTEGRAL      Likelihood = -2.76      Transmembrane      21 - 37 ( 19 - 39)

      ----- Final Results -----
10          bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10129> which encodes amino acid sequence <SEQ ID 10130> was also identified.

15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1608

20 A DNA sequence (GBSx1703) was identified in *S.agalactiae* <SEQ ID 4967> which encodes the amino acid sequence <SEQ ID 4968>. Analysis of this protein sequence reveals the following:

```

      Possible site: 36
      >>> Seems to have no N-terminal signal sequence

25      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.1783(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1609

35 A DNA sequence (GBSx1704) was identified in *S.agalactiae* <SEQ ID 4969> which encodes the amino acid sequence <SEQ ID 4970>. This protein is predicted to be probable 1,4-dihydroxy-2-naphthoate octaprenyltransferase. Analysis of this protein sequence reveals the following:

```

      Possible site: 32
      >>> Seems to have no N-terminal signal sequence

40      INTEGRAL      Likelihood = -8.76      Transmembrane      239 - 255 ( 219 - 260)
          INTEGRAL      Likelihood = -8.33      Transmembrane      221 - 237 ( 219 - 238)
          INTEGRAL      Likelihood = -6.74      Transmembrane      91 - 107 ( 89 - 113)
          INTEGRAL      Likelihood = -6.32      Transmembrane      39 - 55 ( 35 - 59)
          INTEGRAL      Likelihood = -3.77      Transmembrane      111 - 127 ( 111 - 132)
45      INTEGRAL      Likelihood = -2.97      Transmembrane      144 - 160 ( 143 - 161)
          INTEGRAL      Likelihood = -1.28      Transmembrane      275 - 291 ( 275 - 291)
          INTEGRAL      Likelihood = -0.59      Transmembrane      177 - 193 ( 177 - 193)

```

-1792-

----- Final Results -----

bacterial membrane --- Certainty=0.4503(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15875 GB:Z99123 alternate gene name: ipa-6d~similar to
 quinone biosynthesis [Bacillus subtilis]

Identities = 75/290 (25%), Positives = 139/290 (47%), Gaps = 15/290 (5%)

Query: 5 IFLELVEMKAKTASVLPFLIGLCFSAYYNSVHPVYVGLFFVAMFLNMFVDIWNNDY 64
 I +L TAS +P L+G + +Y +++ + F +++ + +++N Y D+
 Sbjct: 21 ILWQLTRPHTLTASFVPVLLGTVLAMFYVKVDLLFLAMLFSCSLWI-QIATNLFNEYDF 79

Query: 65 RNAVDL-DYKNDTNIIGRENLSLRQIEVIMASLVITSSMIGLVLVSVQVGLPLLWMGLFCF 123
 + +D + I R + + I + + + ++G+ + + L +GL
 Sbjct: 80 KRGLDTAESVGIGGAIVRHGMKPKTILQLALASYGIAILLGVICASSSWWLALIGLVGM 139

Query: 124 GIGVLYSFGPRPLSSLPLGEVFSGLTMGFMISLICVYLNTYQNFSDILNLSKIFLISLP 183
 IG LY+ GP P++ P GE+FSG+ MG + LI ++ T D +N+ I LIS+P
 Sbjct: 140 AIGYLYTGGLPIAYTPFGELFSGICMGSVFVLISFFIQT-----DKINMQSI-LISIP 192

Query: 184 NTLWIANLMLANNLCDKEDEKNHRYTLVHYTGIRGGLLLFAISNSIALLAIVFEFLFGL 243
 + + + L+NN+ D EED+K R TL G +G + L A S ++A + +V + G
 Sbjct: 193 IAILVGAINLSNNIRDIEDKKGRKTLAILMGHKGAVTLAASFAVAYIWWVGLVITGA 252

Query: 244 APVTVLLSLLLPFIYKQTKLLWQKQVKRETFVCAVRILALGSATQVLT 293
 A + + L +P + K Q ++ I+A+ S Q T+
 Sbjct: 253 ASPWLFVFLSVKPKVQAVKGFVQNEPMN-----MIVAMKSTAQTNTF 296

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1610

A DNA sequence (GBSx1705) was identified in *S.agalactiae* <SEQ ID 4971> which encodes the amino acid sequence <SEQ ID 4972>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.22 Transmembrane 155 - 171 (154 - 171)

----- Final Results -----

bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15200 GB:Z99120 similar to NADH dehydrogenase [Bacillus subtilis]
 Identities = 178/403 (44%), Positives = 249/403 (61%), Gaps = 7/403 (1%)

Query: 3 EILVLGAGYAGLKAVRNLOKQSG--DFHITLVDMNDYHYEATELHEVAAGSQPKEKITFP 60
 +I++LGAGY GL V L K G D ITLV+ ++YHYE T +HE +AG+ ++ +
 Sbjct: 7 KIVILGAGYGGMLTIVTRLTKYVGPNDADITLVNKHNYHYETTWMEASAGTLHHDRCRYQ 66

Query: 61 IKDVINTNKVNFMQDEVLRVDAENKTVTVKNNGELHYDYVVVALGFVSETFGIKGAMENA 120
 IKDVIN ++VNF+QD V + + K V + N GEL YDY+V+ LG V ETFGIKG E A
 Sbjct: 67 IKDVINQSRVNFVQD TVKAIKIDEKKVLAN-GELQYDYLVLIGLGAVPETFGIKGLKEYA 125

Query: 121 LQMTNISQAENIHNHIVNTMKLYRETKDE--NLLKLLVCGAGFTGIELAGAMVDERPKYA 178
 + NI+ + + HI Y ++ + L ++V GAGFTGIE G + P+

-1793-

Sbjct: 126 FPIANINTSRLLREHIELQFATYNTAEKRPDRLTIVVGAGFTGIEFLGELAARVPELC 185

Query: 179 ALAGVKPEQIEIICVEAATRIILPMFDELAQYGVNLIKDLGINMLGSMIKEIKPGEVVY 238
V + IICVEAA +LP FD EL Y V+ +++ G+ +G+ ++E P V

Sbjct: 186 KEYDVDRLSLVRIICVEAAPTLPFGFDPDELVDYAVHYLEENGVEFKIGTAVQECTPEGVVRV 245

Query: 239 GTSKEDEELKSITAGTIIWTTGVSGSPVMGESGFDQRRGRVMVNSDLRDPKYDNVYVIGD 298
G K+DEE + I + T++W GV G P++ E+GF+ RGRV VN DLR P +DNV+++GD

Sbjct: 246 G--KKDEEPEQIKSQTVVWAAGVRGHPIVEEAGFENMRGRVKVNPDLRAPGHDNVFILGD 303

Query: 299 VSAFMDTESGRPFPTTAQIAITRMGAHVAKNLLHQIKGEATEDFSYSPQGTVASVGNTHGL 358
S FM+ ++ RP+P TAQIA + G VAKNL IKG E+F +GTVAS+G ++

Sbjct: 304 SSLFMNEDTERPYPPTAQIAMQQGITVAKNLGRLIKGELEEFKPDIKGTVASLGEHNAV 363

Query: 359 GVVGKTKIKKYPASVMKKIIMNKS LVD MGGLKELLAKGRFDLY 401
GVV K+K PAS MKK+I N+SL +GGL L KG+F +

Sbjct: 364 GVVYGRKLKGTASFMKKVIDNRSLFMIGGLGLTLKKGKFKFF 406

There is also homology to SEQ ID 4666.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1611

A DNA sequence (GBSx1706) was identified in *S. agalactiae* <SEQ ID 4973> which encodes the amino acid sequence <SEQ ID 4974>. This protein is predicted to be cytochrome d ubiquinol oxidase, subunit I (cydA-1). Analysis of this protein sequence reveals the following:

Possible site: 42
>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -6.64	Transmembrane	19 - 35 (15 - 38)
INTEGRAL	Likelihood = -5.73	Transmembrane	226 - 242 (222 - 244)
INTEGRAL	Likelihood = -4.94	Transmembrane	130 - 146 (126 - 149)
INTEGRAL	Likelihood = -4.83	Transmembrane	429 - 445 (422 - 446)
INTEGRAL	Likelihood = -3.77	Transmembrane	55 - 71 (53 - 74)
INTEGRAL	Likelihood = -3.56	Transmembrane	342 - 358 (340 - 359)
INTEGRAL	Likelihood = -1.06	Transmembrane	89 - 105 (89 - 106)
INTEGRAL	Likelihood = -0.59	Transmembrane	186 - 202 (186 - 202)

----- Final Results -----

bacterial membrane	---	Certainty=0.3654(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15902 GB:Z99123 cytochrome bd ubiquinol oxidase (subunit I)
[Bacillus subtilis]

Identities = 246/470 (52%), Positives = 319/470 (67%), Gaps = 12/470 (2%)

Query: 6 LARFQFAMTTVFHFFVFPFTIGTCLVVAIMETMYVITKNEEYKKLTKFWGNIMLLSFAVG 65
LAR QFA TT+FH FVP +IG +VA+MET+Y++ KNE Y K+ KFWG++ L++FAVG

Sbjct: 6 LARIQFASTTLFHFLFVPMISGLVFMVALMETLYLVKKNELYLKMAKFWGHLFLINFAVG 65

Query: 66 VVTGIIQEFQFGMNWSDYSRFGDIFGAPLAIEALLAFFMESTFLGLWMFTWDNKKISK 125
VVTGI+QEFQFG+NWS DYSRFGD+FGAPLAIEALLAFFMES F+GLW+F WD ++ KK

Sbjct: 66 VVTGILQEFQFGLNWS DYSRFGDVFGAPLAIEALLAFFMESIFIGLWIFGWD--RLPKK 123

Query: 126 LHVTFIWL VVFGSLMSAMWILTANSFMQHPVGYEVVNGRAQMTDFLALVKNPQFFYEFT 185
+H IWL V FG++MS+ WILTANSFMQ PVG+ + NGRA+M DF AL+ NPQ + EF H

Sbjct: 124 IHALCIWLVSFGTIMSSFWILTANSFMQEPVGFITKNGRAEMNDFGALITNPQLWVEFP 183

Query: 186 VIFGAIMGCTVVAGMSAFRLKSEQLKDTTVELYKKSVRIGLVALLGSISVMGVGDLQ 245
VIFGA+ G +AG+SAF+LLK ++ V +K+S ++ ++V L + V G +Q

-1794-

Sbjct: 184 VIFGALATGAFFIAGVSAFKLLKKKE-----VPFFKQSFKLAMIVGLCAGLGVGLSGHMQ 238

Query: 246 MKALIHDPMPKFAAMEGDYEDSGDPAAWSVVAWANEAEHKQVFGIKIPYMLSILSYGKPS 305
+ L+ QPMK AA EG +EDSGDPAAW+ A + K IK+PY LS L+Y K S

5 Sbjct: 239 AEHLMESQPMKMAASEGLWEDSGDPAAWTAFATIDTKNEKSSNEIKVPYALSILAYQKFS 298

Query: 306 GSVKGMTANKELVAKYGKDNYPVNVNLLFYGFRTMAAMGTAIMGVSVLGLFLTRKKKPI 365
GSVKGM T E YGK +Y P V F+ FR M G ++ ++ GL+L R+KK

10 Sbjct: 299 GSVKGMKTLQAEYEKIYKGDYIPPVKTTFSFRIMVGAGVVMILAAALGGLWLNRRKK-- 356

Query: 366 LYKHKMWLWIVALTTTFAPFLANTFGWIVTEQGRYPWTIVYGLFKIKDSVSPNVSVASLFVS 425
L KW L I+ PFLAN+ GWI+TE GR PWTIV GL SVSPNV+ SL S

15 Sbjct: 357 LENSKWYLRIMIALISFPFLANSAGWIMTEIGRQPWTVMGLMTTAQSVSPNVTAGSLLFS 416

Query: 426 NTVYFLLFGGLAVMMISLTIRELKKGPEYEDEHGHGAYTSIDPFEEGAY 475
+ +++ L +++ L IRE+KKG E+++ HH S DPF + Y

20 Sbjct: 417 IIAFGVMYMLGALLVFLFIREIKKGAEDN--HHDVPVSTDPFSEQEVY 463

No corresponding DNA sequence was identified in *S.pyogenes*.

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1612

A DNA sequence (GBSx1707) was identified in *S.agalactiae* <SEQ ID 4975> which encodes the amino acid sequence <SEQ ID 4976>. This protein is predicted to be cytochrome oxidase subunit II (cydB-1).

25 Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have an uncleavable N-term signal seq

30	INTEGRAL	Likelihood = -14.49	Transmembrane	226 - 242 (220 - 250)
	INTEGRAL	Likelihood = -8.12	Transmembrane	254 - 270 (250 - 282)
	INTEGRAL	Likelihood = -7.64	Transmembrane	198 - 214 (196 - 218)
	INTEGRAL	Likelihood = -6.95	Transmembrane	85 - 101 (76 - 103)
	INTEGRAL	Likelihood = -6.74	Transmembrane	6 - 22 (1 - 27)
	INTEGRAL	Likelihood = -6.16	Transmembrane	300 - 316 (298 - 322)
	INTEGRAL	Likelihood = -5.36	Transmembrane	119 - 135 (117 - 143)
35	INTEGRAL	Likelihood = -4.04	Transmembrane	159 - 175 (155 - 178)

----- Final Results -----

	bacterial membrane	--- Certainty=0.6795(Affirmative) < succ>
40	bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15901 GB:Z99123 cytochrome bd ubiquinol oxidase (subunit II)
[Bacillus subtilis]

45 Identities = 158/331 (47%), Positives = 223/331 (66%), Gaps = 1/331 (0%)

Query: 1 MSALQFFWFFLIGLLFSGGFFLEGDFGVGMVAVQTLTHNEHEKDQVETIGPVDGNEVW 60
M++L WF L+ +LF GFFLEGDFGVGMA + L HNE E+ ++ TIGP WD NEVW

50 Sbjct: 1 MASLHDLWFLVAVLVFGFFLEGDFGVGMATRFLGHNELERRVLINTIGPFDANEVW 60

Query: 61 LLTGGGAMFASFPYWYASLFSGYLLILLTILFGLIIRGVSFETRHKVPAEK-KQFWNWTL 119
LLTG GA+FA+FP WYAA+ SGYY+ + +L L+ RGV+FEFR KV K + W+W +

Sbjct: 61 LLTGAGAIFAAFPNWYATMLSGYYIPFVIVLLALMGRGVAFEFRGKVDHLKWKVVDWV 120

55 Query: 120 TIGSAIVPFFFGIMFISLIQGMPLDASGNLSAQFSDFYFNIFSLVGGVAMVLLAYLHGLNY 179
GS I PF G++F +L +GMP+DA N+ A SDY N++S++GGV + LL + HGL +

Sbjct: 121 FFGSLIPPFVLGVLFITLFRGMPIDAMNTHAHVSDYINVYSILGGVTVTLLCFQHGLMF 180

60 Query: 180 IALKTEGPPIRERARNYAQLLYWVLYLGLALFAVLLYFKTDFFSNHPITVTIMVLVIVVLA 239
I L+T G ++ RAR AQ + V+++ + FA L ++TD F+ +T + ++IV+

-1795-

Sbjct: 181 ITLRTIGDILQNRARKMAQKIMGVVFVAFLAFAALSAYQTMFTRRGEITIPLAVLIVICF 240

Query: 240 VLAHASTFKGAEMTAFLASGLSLVSVVLLFQGLFPRVMISSISPKYDLLIQNASSTPYT 299

+LA K + F +G L V ++F LFPRVM+SS+ YDL + NASS Y+

Sbjct: 241 MAAVFIRKKKDGWTFGMTGAGLALTVMIFISLFPRVMVSSLHSAYDLTVANASSGDYS 300

Query: 300 LKVMISIVAITLVPFVLAYTAWAYYIFRKRIT 330

LKVMSI A+TL+PFV+ W+YY+FRKR++

Sbjct: 301 LKVMISIAALTLLPFVIGSQIWSYVFRKRVS 331

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1613

- 15 A DNA sequence (GBSx1708) was identified in *S.agalactiae* <SEQ ID 4977> which encodes the amino acid sequence <SEQ ID 4978>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 25 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1614

- 30 A DNA sequence (GBSx1709) was identified in *S.agalactiae* <SEQ ID 4979> which encodes the amino acid sequence <SEQ ID 4980>. This protein is predicted to be transport ATP-binding protein cydc (cydD). Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -16.82 Transmembrane 158 - 174 (144 - 182)

INTEGRAL Likelihood = -6.48 Transmembrane 15 - 31 (14 - 34)

INTEGRAL Likelihood = -5.31 Transmembrane 243 - 259 (238 - 266)

INTEGRAL Likelihood = -2.55 Transmembrane 136 - 152 (134 - 152)

INTEGRAL Likelihood = -0.48 Transmembrane 263 - 279 (263 - 279)

----- Final Results -----

bacterial membrane --- Certainty=0.7729(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15900 GB:Z99123 ABC membrane transporter (ATP-binding protein) [Bacillus subtilis]

Identities = 279/569 (49%), Positives = 401/569 (70%), Gaps = 6/569 (1%)

Query: 2 LDKAVMRLSGIHKLGLLGLAGLDVLQAIIFIIGQAYYLSLSITGLWEGQKLSSQTVYILLFM 61

+ K + R G+ ++L L+ L ++Q II QA +LS ++TGL+ G+ ++S I F+

Sbjct: 1 MGKDLFRYKGMKRILTLITCLTLIQTAAIMQAEWLSEAVTGLENGKGITSLLPVIGFFL 60

-1796-

Query: 62 VSYLGRHVIDYIKNRKLDFFSTAQSSILRRQLLDKLFDLGPKVVQEQTGNVVTMALDGV 121
 +++++ RH + + + + ++ + LR+ LD+LF LGP+ +++GTG +VT+A++G+
 Sbjct: 61 IAFIARHGMTVARQKIVYQYAARTGADLRKSFLLDQLFRLGPRFAKKEGTGMVTLAMEGI 120

Query: 122 SLVENYLRLVLNKNMINSIIPWILAYIFYLDIESGAILLIVFPLIIFMIILGYAAQAK 181
 S YL L L KM++M+I+P ++ Y+F+ D S IL+ P++IIFMI+LG AQ K
 Sbjct: 121 SQFRRYLELFLPKMVSMAIVPAAVVIYVFFQDRTSAILVAAMPILIIIFMILLGLVAQRK 180

Query: 182 ADKQYESYQVLSNHFLDSLRLGIDTLKYFGLSKRYGKSIYQTSSESFRKATMSTLKIGILST 241
 AD+Q++SYQ LSNHF+DSLRLG++TL++ GLSK + K+I+ SE +RKATMSTL++ LS+
 Sbjct: 181 ADRQWKSQRLSNHFVDSLRLGLETLRLFLGLSKSHSKNIFYVSERYRKATMSTLRVAFLLSS 240

Query: 242 FALDFFFTLSIAIVAVFLGLRLRLNEQIYLLPALTILILSPEYFLPVRDFSSDYHATLDGK 301
 FALDFFFT LS+A VAVFLGLRL++ I L PALT LIL+PEYFLPVR+ +DYHATL+G+
 Sbjct: 241 FALDFFFTLSVATVAVFLGLRLIDGDILLGPALTALILAPEYFLPVREVGN DYHATLNGQ 300

Query: 302 NAFQAIQKVLNKTGIKGE-QLVIDDWSKESRLDLENIAIAYDQKRVVEDVTLRFRGHQKV 360
 A + IQ++L++ G K E L ++ WS + L L +++ R V D+ L F+G +K+
 Sbjct: 301 EAGKTIQEILSQPGFKEETPLQLEAWSQDELKLSGVSVG----RSVSDIHL SFGKGGKKI 356

Query: 361 ALVGVSGSGKSSLINLLSGFLGPDNGSLKVDGREVTNLDQEDWHKQMIYIPQTPYVFEMS 420
 ++G SG+GKS+LI++L GFL PD G ++V+G ++L W K ++YIPQ PY+F+ +
 Sbjct: 357 GIIGASGAGKSTLIDILGGFLBPDGGMIEVNGTSRSHLQDGSWQKNLLYIPQHPYIFDDT 416

Query: 421 LRDNITFYTPNASDEEVVRAIHMVGLDLSLSELPDGLETRIGNGARPLSGGQAQRALAR 480
 L +NI FY P+AS E+ RA GL L++ LPDGLE RIG G R LSGGQAQR+ALAR
 Sbjct: 417 LGNNIRFYHPSASAEDTTRAAASAGLTENVNLPDGLEGRIGEGGRALSGGQAQRVALAR 476

Query: 481 AFLDQNRIMVFEPTAHLDDIETELELEKEKMLPLMSDRLVIFATHRLHLWLNQMDVIVVME 540
 AFL NR I++ DEPTAHLDDIETE E+KE ML L D+LV ATHRLHW+ MD I+V++
 Sbjct: 477 AFLG-NRPILLDEPTAHLDDIETEYEIKETMLDLFEDKLVLATHRLHWMLDMDEIIVLD 535

Query: 541 KGRVAEVSQYQELLAKKGYLYQLKHAMGG 569
 GRVAE+G++ ELL K G +L A G
 Sbjct: 536 GGRVAEIGTHNELLEKNGVYTKLVKAQLG 564

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4981> which encodes the amino acid sequence <SEQ ID 4982>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.61	Transmembrane	159 - 175 (154 - 190)
INTEGRAL	Likelihood = -10.03	Transmembrane	70 - 86 (63 - 91)
INTEGRAL	Likelihood = -3.03	Transmembrane	282 - 298 (282 - 301)
INTEGRAL	Likelihood = -1.44	Transmembrane	261 - 277 (260 - 278)

----- Final Results -----
 bacterial membrane --- Certainty=0.5246(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC22320 GB:U32749 ATP-binding transport protein (cydD)
 [Haemophilus influenzae Rd]
 Identities = 167/544 (30%), Positives = 279/544 (50%), Gaps = 15/544 (2%)

Query: 46 MISFYLIATFSTFILGHAIALGRLAGLLLLLVGVFLAILGK---QLQGIASQFARDS 102
 + S+ L A F L A+ LG + L L A GK Q AS +
 Sbjct: 17 VFSYILQAAYFHELSLLSAVILGIVLIAALALR-----AFAGKKSQVQASYFASTKVKHE 70

Query: 103 LKQSFFFEAFIDLDGQFDAHASDADILTLASQGIDSLDTYYGYL-SLSMRTKWNCTTIMI 161
 L+ + + S + I+ +AS+G++ L+ Y+G YL L T
 Sbjct: 71 LRSLIYRKLASMPINQVNQOSTSSIIQVASEGVEQLEIFYGRYLPQLFYSLLAFLTLFAF 130

Query: 162 LVFLIYPLAGLVFLGVLPPLIPLSIVAMQKRSQPNMSHYWSSYMDVGNLFMDDLKGLNPLY 221

-1797-

L+F + A ++ L +PLIP+SI+A+ K ++ ++ YWS Y+ +G+ F+D+L+GL TL
 Sbjct: 131 LIFFSFKTA-IILLICVPLIPMSIIAVNKIAKLLAKYWSIYVGLGSSFLDNLQGLITLK 189
 Query: 222 SYQATERYEQEFSGKAEQFRKATMSLLGFQLQAVGYMDAVMYLGIGLSGFLAVQALATGQ 281
 YQ + +AE FRK TM +L QL +V MD + Y G + A+ Q
 Sbjct: 190 IYQDDAYKAKAMDKEAEHFRKTMKVLTMQLNSVSLMDLLAYGGAIGILTALLQFQNAQ 249
 Query: 282 LSFFNFFLLIATEFFFTPIREQGYGMHLVMMNTKMADRIFSFLDS-VPARKDNKSKTAI 340
 LS + F+L+++EFF P+R G H+ M +D+IF+ LD+ V ++ A
 Sbjct: 250 LSVLGVILFILLSSEFFIPLRLLGSGFFHVAMNGKAASDKIFTLLDTPVETQQSAVDFEAK 309
 Query: 341 NFNQIDIQNISLAY-EKKTIVLSGVITMTLTKGQITAIAGVSGQKTSIAQLLLKRSATTG 399
 N Q++I+++ +Y E+K ++G+ +++ QL+ G SG GK++L LL+ A G
 Sbjct: 310 NNVQVEIKDLHFSYSEKPAITGLNLSILPNQLSVFVGKSGCGKSTLVSLMGMFNKAQQG 369
 Query: 400 HILFDGLSDNLSQETINQQVLYVSDQSTLLNRSIYDNLRLA-ANLSKKEILDWIDQHGL 458
 ILF+G ++ N+ + + Q+V VS S + ++ +N+ +A + + ++I ++Q L
 Sbjct: 370 EILFNGQNALNIDRTSFYQKVSLSVSHSSYVFKGTLRENMTMAKIDATDEQIYACLEQVNL 429
 Query: 459 LSFINWLPDGLDITVGENGNLSPGQKQOVICARALLSKRSIYIFDEATSSSLDAENERII 518
 F+ GLD + G LS GQ Q++ ARALL LYIFDEATS++D E+E II
 Sbjct: 430 AQFVR-DNGGLDMQLLSRGANLSSGQIQRLALARALLHNAELYIFDEATSNIDVESEEII 488
 Query: 519 DNLITRLAKTAIVIVITHKMSRLKGANQVFLNTGQPACLGKPCDLYRDQPTYRHLVDTQ 578
 I + + +++I+H+++ A+ + L+ G+ G +L Q Y + Q
 Sbjct: 489 LQFIQQFQKQKTIVMISHRLANAVNADCINVLDDQGLIEQGT HKELMEKQGAYAEMFQQQ 548
 Query: 579 ARLE 582
 LE
 Sbjct: 549 KDLE 552

An alignment of the GAS and GBS proteins is shown below.

Identities = 143/552 (25%), Positives = 260/552 (46%), Gaps = 12/552 (2%)

Query: 1 MLDKAVMRLSGIHKLGLLAGLDVLQAIFFIIGQAYYLSLSITGLWEGQKLSSQTVYILLF 60
 +L + R++ LL + A L LQ + + Y ++ + + G ++ + LL
 Sbjct: 16 LLKRLRERIAPKRYLLYVSAFLSWLQFVMRMISFYLIAKTFSTFILGHAIAGRLAGLLL 75
 Query: 61 MVSYLGRHVIDYIKNRKLDLDFSTAQSSLLRRQLLDKLFDLGPKVVQEQTGNVVTMALDG 120
 +++ +G V+ + + S L++ + DL + +++T+A G
 Sbjct: 76 LLNVVG-FVLAILGKQLQGIASQFARDSLKQSFFFAFIDLGGQFDAHASADILTLASQG 134
 Query: 121 VSLVENYLRVLNKMINSIIPWIIAYIFYLDIESGAILLIVFPLIIFMIILGYAAQA 180
 + ++ Y L+ + I+ +F + +G + L V PLI + ++ + +Q
 Sbjct: 135 IDSLDTYYGYLSLSMRKWNCTTIMILVFLIYPLAGLVFLGVLPLIPLSIVAMQKRSQP 194
 Query: 181 KADKQYESYQVLSNHFLDSLGRIDTLKYFGLSKRYGKSIYQTSSESFRKATMSTLKIGILS 240
 + SY + N F+D L+G++TL + ++RY + +E FRKATMS L + +
 Sbjct: 195 NMSHYWSSYMDVGNLFMDLKLGLNTLYSYQATERYEQEFSGKAEQFRKATMSLLGFQLQA 254
 Query: 241 TFALDFFTTLSIAIVAVFLGLRLINEQIYLLPALITILLSPEYFLPVRDFSSDYHATLDG 300
 +D L I + L Q+ L L+++ E+F P+R+ H +
 Sbjct: 255 VGMDAVMYLGIGLSGFLAVQALATGQLSFFNFFLLIATEFFFTPIREQGYGMHLVMMN 314
 Query: 301 KNAFQAIQKVLNKTGIGKEQLVIDDWSKE----SRLDLENIAIAYDQKRVVEDVTLRFRG 356
 I L+ + D+ SK +++D++NI++AY++K V+ VT+
 Sbjct: 315 TKMADRIFSFLDSVPARK-----DNKSKTAINFNQIDIQNISLAYEKKTVLSGVITMTLTK 369
 Query: 357 HQKVALVGVSQKSSLLINLLSGFLGPDNGSLKVDGREVINLDQEDWHKQMIYIPQTPYV 416
 Q A+ GVSG GK+SL LL G + DG + NL QE ++Q++Y+ +
 Sbjct: 370 GQLTAIAGVSGQKTSIAQLLLKRSATTGHILFDGLSDNLSQETINQQVLYVSDQSTL 429
 Query: 417 FEMSLRDNITFYTPNASDEEVVRAIHMVGLDSLSELDPGLETRIGNGARPLSGGQAQRI 476
 S+ DN+ N S +E++ I GL S ++ LPDGL+T +G LS GQ Q++
 Sbjct: 430 LNRSIYDNLRL-AANLSKKEILDWIDQHGLLSFINWLPDGLDITVGENGNLSPGQKQOV 488
 Query: 477 ALARAFLDQNRIMVFDEPTAHLDIETELKEKMLPLMSDRLVIFATHRLHWNQMDVI 536

-1798-

ARA L + R + +FDE T+ LD E E + + L +VI TH++ L + +
 Sbjct: 489 ICARALLSK-RSLYIFDEATSSLDAENERIIDNLITRLAKTAIVIVITHKMSRLKGANQV 547

Query: 537 VVMEKGRVAEVG 548
 + + G+ A +G
 Sbjct: 548 LFLNTGQPACLG 559

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 1615

A DNA sequence (GBSx1710) was identified in *S.agalactiae* <SEQ ID 4983> which encodes the amino acid sequence <SEQ ID 4984>. This protein is predicted to be transport ATP-binding protein cydd (cydC). Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -12.84	Transmembrane	260 - 276 (258 - 284)
INTEGRAL	Likelihood = -9.34	Transmembrane	172 - 188 (147 - 199)
INTEGRAL	Likelihood = -6.53	Transmembrane	150 - 166 (147 - 171)
INTEGRAL	Likelihood = -6.05	Transmembrane	31 - 47 (29 - 52)
INTEGRAL	Likelihood = -3.35	Transmembrane	68 - 84 (67 - 84)
INTEGRAL	Likelihood = -1.17	Transmembrane	293 - 309 (292 - 310)
INTEGRAL	Likelihood = -0.69	Transmembrane	494 - 510 (493 - 510)

----- Final Results -----

bacterial membrane --- Certainty=0.6137(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10127> which encodes amino acid sequence <SEQ ID 10128> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15899 GB:Z99123 ABC membrane transporter (ATP-binding protein) [Bacillus subtilis]
 Identities = 262/573 (45%), Positives = 389/573 (67%), Gaps = 14/573 (2%)

Query: 16 LKTDQWIKPFFKQYKVSIVIALFLGFMTFFSASALMFNSGYLISKASALPSNILLVYVPI 75
 +K ++WI P+ KQ V+ +FLG +T FSA+ LMF SG+LISK+A+ P NILL+YVPI
 Sbjct: 1 MKKEEWILPYIKQARLFVLVIFLGAVTIFSAFLMFTSGFLISKAATRPENILLIYVPI 60

Query: 76 VLTRAFGIGRPVFRYIERLTSHNVLRMTSQLRLKLYHSLESNAIFMKRDFRLGDVMGLL 135
 V R FGI R V RY+ERL H+ +L++ S +R++LY+ LE A+ ++ FR GD++G+L
 Sbjct: 61 VAVRTFGIARSVSRVYVERLVGHIIILKIVSDMRVRLYNMLEPGALMLRSRFRGTGDMGLIL 120

Query: 136 AEDINYLQNLVLRITFPTIIAWILYSFIIATGFFSLWFLMMLLYLAIMIFLFLWSIL 195
 +EDI +LQ+ +L+TIFP I A +LY+ +IA GFFS FA+++ LYL +++ LFP+ S+L
 Sbjct: 121 SEDIEHLQDAFLKTIFFPAISALLLYAVSVIALGFFSWPFAILLALYLFVILVFLFPVVSLL 180

Query: 196 ANGARQTRKELKNHLYTDLTDNLVIGISDWIFSQRGQEXVALHERSESELMVQKKIRSF 255
 A+ + K +N LY+ LTD V+G+SDW+FS R ++ +E+ E + +++K + F
 Sbjct: 181 VTRAKNAKLKSGRNVLYSRLTDVAMGVSDWMFSGRRHAFIDAYEKEERDWFELERKKQRF 240

Query: 256 DNRRALIVELVFGFLAILVLIWASNQFIGHRGGEA--NWIAAFVLTVFPLSEAFAGLSAA 313
 R + + L +L++ W + Q GE IAAFVL VFPL+EAF LS A
 Sbjct: 241 TRWRDFAAQCLVAGLILLMLFWTAGQ---QADGELAKTMIAAFVLVVFPLTEAFLPLSDA 297

Query: 314 AQETNKYSDSIHRLN-----ELSETYFETTONQLENKPYDFSVKNLSFQYKPKQEKWVLH 367
 E Y DSI R+N E S+T E+ L + + ++++F Y + VLH
 Sbjct: 298 LGEVPGYQDSIRRMNNVAPQPEASQT--ESGDQILDLDQVTLAFRDVTFSDNNSQ-VLH 354

Query: 368 HLDLDIKEGEKIAILGRSGSGKSTLASLLRGDLKASQGEITLGDADVSI V GDCISNYIGV 427
+ +++GEK+A+LGRSGSGKST +L+ G LK G +TL + +++ D I++ + V
Sbjct: 355 NFSFTLRQGEKMA LLGRSGSGKSTSLALIEGALKPDSGSVTLNGVETALLKDQIADAVAV 414

5 Query: 428 IQQAPYLFNTTLLNNIRIGNQDASEEDVWKVLERVGLKEMVTDLS DGLYTMVDEAGLRFS 487
+ Q P+LF+T++LNNIR+GN +AS+EDV + ++V L + + L DG +T V E G+RFS
Sbjct: 415 LNQKPHLFDTSILNNIRLNGEASDEDVRRAAQVKLHDYIESLPDGYHTSVQETGIRFS 474

10 Query: 488 GGERHRIALARILLKDVPIVILDEPTVGLDPITEQALLRVFMKELEGKTLVWITHHLKGI 547
GGER RIALARILL+D PI+ILDEPTVGLDPITE+ L+ + L+GKT++WITHHL G+
Sbjct: 475 GGERQRIALARILLQDTPIIILDEPTVGLDPITERELMETVFVFLKGGKTTILWITHHLAGV 534

Query: 548 EHADRILFIENGQLELEGSPQELSQSSQRYRQL 580
E AD+I+F+ENG+ E+EG+ +EL +++RYR+L
Sbjct: 535 EAADKIVFLENGKTEMEGTHEELAANERYRRL 567

15

```

Lipop: Possible site: -1   Crend: 8
20  McG: Discrim Score:    -15.90
    GvH: Signal Score (-7.5): 1.97
        Possible site: 49
    >>> Seems to have no N-terminal signal sequence
ALOM program    count: 7 value: -12.84 threshold:  0.0
25  INTEGRAL    Likelihood ==-12.84   Transmembrane  260 - 276 ( 258 - 284)
    INTEGRAL    Likelihood = -9.34    Transmembrane  172 - 188 ( 147 - 199)
    INTEGRAL    Likelihood = -6.53    Transmembrane  150 - 166 ( 147 - 171)
    INTEGRAL    Likelihood = -6.05    Transmembrane   31 - 47 ( 29 - 52)
    INTEGRAL    Likelihood = -3.35    Transmembrane   68 - 84 ( 67 - 84)
30  INTEGRAL    Likelihood = -1.17    Transmembrane  293 - 309 ( 292 - 310)
    INTEGRAL    Likelihood = -0.69    Transmembrane  494 - 510 ( 493 - 510)
    PERIPHERAL  Likelihood =  3.29      412
    modified ALOM score:   3.07

35  *** Reasoning Step: 3

----- Final Results -----
        bacterial membrane --- Certainty=0.6137(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
40  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

```

ORF00997(346 - 2052 of 2364)
EGAD|98910|BS3866(1 - 571 of 575) transport ATP-binding protein cydD {Bacillus subtilis}
45 OMNI|NT01BS4517 ABC transporter CydC, putative SP|P94367|CYDD_BACSU TRANSPORT ATP-BINDING
PROTEIN CYDD. GP|1783253|dbj|BAA11730.1||D83026 homologous to many ATP-binding transport
proteins; hypothetical {Bacillus subtilis} GP|2636408|emb|CAB15899.1||Z99123 ABC membrane
transporter (ATP-binding protein) {Bacillus subtilis} PIR|D69611|D69611 ABC transporter
required for expression of cytochrome bd (ATP-) cydD - Bacillus subtilis
50 %Match = 31.9
%Identity = 45.2 %Similarity = 69.1
Matches = 257 Mismatches = 172 Conservative Sub.s = 136

300          330          360          390          420          450          480          510
55 LKKDISIN*SMLWEEMMFKIPLFKELKTDQWIKPFFKQYKVSLLVIALFLGFMTFFSASALMFNSGYLISKASLSPSNILL
      :|::||| |:|| :|: :||| :| |||: ||| ||:||||:|: | ||||
      MKKEEWILPYIKQARLFVLVIFLGAVTIFSAAFMFTSGFLISKAAATRPENILL
      10          20          30          40          50

540          570          600          630          660          690          720          750
60 VYVPIVLTRAFGIGRPVFRYIERLTSHNWVLRMTSQLRLKLYHSLESNAIFMKRDFRLGDVGMGLAEDINYLQNLRLRTI
      :||||| | ||| | | ||:||| |: :::: | :::||: || |: : : || ||:|:|:| | |:|: :|:|
      IYVPIAVARTFGIARSVSRYVERLVGHIIILKIVSDMRVRLYNMLEPGALMLRSRFRGTGDMILGILSEDIHQLQDAFLKTI
      70          80          90          100          110          120          130

```

-1800-

```

5      780      810      840      870      900      930      960      990
FPTIIAWILYSFIIATGFFSLWFMMLLYLAIMIFLFPFLWSILANGARQTRKELKNHLYTDLTDNVLGISDWIFSQR
|| | | :||: :|| |||| ||:: ||| :: |||: |:| |: : | :| ||: ||| |:|:||||:| |
FPAISALLLYAVSVIALGFFSWPFAILLALYLFLVLVLPVVSLLVTRAKNAKLKSGRNVLYSRLITDAVMGVSDWMFSGR

10      1020     1050     1080     1110     1140     1194     1224
GQEVVALHERSESELMAVQKKIRSFDRRALIVELVFGFLAILVLIWASNQFIGHRGGE--ANWIAAFVLTVPFLSEAF
: :: :|: | : ::| : | | : : : | :||: | : | : || ||||| ||||:| |
RHAFIDAYEKEERDWFELERKKQRFTRWRDFAAQCLVAGLILMLFWTAGQ---QADGELAKTMIAAFVLVFPFLTEAFL
230      240      250      260      270      280      290

15      1254     1284     1302     1332     1362     1392     1422     1452
GLSAAAEQETNKYSDSIHRNLNELS----ETVFETQNLQPNKPYDFSVMKLSFQYKPKQEKWVLHHLDLDIKEGEKIAILGR
|| | | | ||| |:| :: : | : | : : ::| | ||:: : ::|||:| |
PLSDALGEVPGYQDSIRRMNNVAPQFEASQTESGDQILDLDQDVTILAFRDVTFYSY-DNSSQVLHNFSTLRQGEKMALLGR
310      320      330      340      350      360      370

20      1482     1512     1542     1572     1602     1632     1662     1692
SGSGKSTLASLLRGDLKASQGEITLGDADSVISVGDICISNYIGVIQQAPYXFNTLLNTFRIGNQDASEEDVWKVLERVGL
||||| | :|: | | | :|| : :: | |:: : |: | |: |:|:| | |:| :|:| | : :| |
SGSGKSTSLALIEGALKPDGSGSVTLNGVETALLKDQIADAVAVLNQKPHLFDTSILNNIRLNGGEASDEDVRRRAKQVKL
390      400      410      420      430      440      450

25      1722     1752     1782     1812     1842     1872     1902     1932
KEMVTDLSGGLYTMVDEAGLRFSGGERHRIALARILLKDVPIVILDEPTVGLDPITEQALLRVFMKELEGKTLVWITHL
: : | | | :| | | :||| |||:| | | | | | | | | | | | | | | | : : | :|:| | | | | | | |
HDYIESLPDGYHTSVQETGIRFSGGERQRIALARILLQDTPIIILDEPTVGLDPITERELMETVFEVLKGTKITLWITHL
470      480      490      500      510      520      530

30      1962     1992     2022     2052     2082     2112     2142     2172
KGIEHADRIIFIENGQLELEGSPQELSQQSQRRLKASDDGDL**LIGAINK***KNIP*LLF*HCGMFFYYLNF*F*K
|:| ||:|:|:|:| |:|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGVEAADKIVFLENGKTEMEGTHEELLAANERYRRLYLHLDVVPVK
550      560      570

```

There is also homology to SEQ ID 478.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1616

A DNA sequence (GBSx1711) was identified in *S. agalactiae* <SEQ ID 4987> which encodes the amino acid sequence <SEQ ID 4988>. This protein is predicted to be spore germination protein C3 (ispB). Analysis of this protein sequence reveals the following:

```

45      Possible site: 45
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -1.06      Transmembrane 111 - 127 ( 111 - 128)

----- Final Results -----
      bacterial membrane --- Certainty=0.1426(Affirmative) < succ>
50      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

55      >GP:CAB14190 GB:Z99115 heptaprenyl diphosphate synthase component II
      [Bacillus subtilis]
      Identities = 101/318 (31%), Positives = 184/318 (57%), Gaps = 5/318 (1%)

Query: 8      YPELKKNIDETNQLIQERIQVRNKDIEAALSQLTAAAGKQLRPAFFYLFSQLGNKENQDT 67
      Y L +ID + +++ ++ + A L AGGK++RP F L G+ D
60      Sbjct: 35      YSFLNDDIDVIERELEQTVRSDYPLLSEAGLHLLQAGGKRIRPVFVLLSGMFGD---YDI 91

```

-1801-

5 Query: 68 QQLKKIAASLEILHVATLIHDDVIDDSPLRRGNMTIQSKFGKDIAVYTGDLLFTVFFDLI 127
 ++K +A +LE++H+A+L+HDDVIDD+ LRRG TI++K+ IA+YTGd + +++
 Sbjct: 92 NKIKYVAVTLEMIHMASLVHDDVIDDAELRRGKPTIKAKWDNRIAMYTGDYMLAGSLEMM 151

10 Query: 128 LESMADTPFMRINAKSMRKILMGELDQMHLRYNQOQGIHHYLRAISGKTAELFKLASKEG 187
 + + RI +++++ ++ +GE++Q+ +YN +Q + YLR I KTA L ++ + G
 Sbjct: 152 TR-INEPKAHRILSQTIVEVCLGEIEQIKDKYNMEQNLRTYLRRIKRTALLIAVSCQLG 210

15 Query: 248 AIEENPDIFKPILDKKTDMATEDMEKIAYLVSHRGVDKARHLARKFTEKAISDINKLPQ 307
 A GA++++ + G+ +GM++QI+DDILD+T+ ++ KPV DL QG +LP+L
 Sbjct: 211 AIASGADEKIKALYWFYVGMYSYQIIDILDFTSTEEELGKPVGGDLQGNVTLPLVLY 270

20 Query: 308 SSAKKQLQLTNYLLKRK 325
 A+ L + Y+ KRK
 Sbjct: 330 GRARSSLAIAKYIGKRK 347

There is also homology to SEQ ID 284. An alignment of the GAS and GBS proteins is shown below:

Identities = 65/227 (28%), Positives = 98/227 (42%), Gaps = 9/227 (3%)

25 Query: 43 AGGKQLRPAFFYLFSQLGNKENQDTQQLKKIAASLEILHVATLIHDDV--IDDSPLRRGN 100
 +GGK++RP + Q+ +AA+LE++H +LIHDD+ +D+ RRG
 Sbjct: 36 SGGKRIRPLILLEMIEGFGVSLQNAHF--DLAAALEMIHTGSLIHDDLPAMDNDYRRGR 93

30 Query: 101 MTIQSKFGKDIAVYTGDLLFTVFFDLILESM--ADTPFMRINAKSMRKILMGELDQMHLR 158
 +T +FG+ A+ GD LF F LI ++ ++ I S+ G + L
 Sbjct: 94 LTNHKQFGEATAILAGDSLFLDPGLIAQAEINSEVKVALIQELSLASGTFGMVGGQMLD 153

35 Query: 159 Y---NQOQGIHHYLRAISGKTAELFKLASKEGAYFGGAEKEVVRLAGHIGFNIGMTFQIL 215
 NQ + KT +L K A V + G IG FQI
 Sbjct: 154 MKGENQALSLPQLSLIHLNKTGKLLAFPFKAALITEQAMTVRQOLEQAGMLIGHAFQIR 213

40 Query: 216 DDILDYTADKKTFNKPVLEDLAQGIYSLPILLAIENPDIFKPILDK 262
 DDILD TA + K +DL + P LL +E + + LD+
 Sbjct: 214 DDILDVTASFEDLGKTPKKDLFAEKATYPSSLGLEASYQLLTESLDQ 260

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1617

45 A DNA sequence (GBSx1712) was identified in *Sagalactiae* <SEQ ID 4989> which encodes the amino acid sequence <SEQ ID 4990>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3995(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:AAA25232 GB:M58315 dipeptidyl peptidase IV [Lactococcus lactis]
 Identities = 385/767 (50%), Positives = 504/767 (65%), Gaps = 21/767 (2%)

60 Query: 1 MRYNQFSYIPTKPNFAFEELKGLGFPLNKKNSDKANLEAFRLHSFNLQTDTDYALSLIV 60
 MR+N FS + +E EL LGF + +K L+ FL S + TD L
 Sbjct: 1 MRFNHFISIVDKNFDEQLAELDQLGFRWSVFWEKKILKDFLIQSPSDMTD-----LQA 53

-1802-

Query: 61 DAKTDALTFKSNSDLTLENLQWIYLLQLLGFIPFVDFKDPKAF-----LQDINFPVSY 113
 A+ D + F KS+ +L E I LQLL F+P DF+ KAF L I ++
 Sbjct: 54 TAELEDVIEFLKSSIELDWEIFWNIALQLLDVFPNFDGFEIGKAFYAKNSNLPQIEAEMTT 113

5 Query: 114 DNIFQSLHLLACRGKSGNTLIDQLVADGLLHADNHYHFFNGKSLATFTNTNQLIREVVYV 173
 +NI + ++LL R K+G L++ V++GLL DNHYHFFN KSLATF+++ L REV++V
 Sbjct: 114 ENIISAFYLLCTRRKNGMILVEHWVSEGLPLDNHYHFFNDKSLATFDSSLLEREVLWV 173

10 Query: 174 ETSLDTMSSGEHDLVKVNIIRPTTEHTIPTMTASPYHQGINDDPAADQKTYQMEGALAVK 233
 E+ +D+ GE+DL+K+ IIRP + +P +MTASPYH GIND A D + M L K
 Sbjct: 174 ESPVDSEQRGENDLIKIQLIRPKSTKLPVVMASPYHLGINDKANDLALHDMNVELEEK 233

15 Query: 234 QPKHIQVDTKPFKEEVKHPKSLPI-SPATESFTHIDSYSLNDYFLSRGFANIYVSGVGTA 292
 I V+ K ++ +LPI A FTH +YSLNDYFL+RGFA+IYV+GVGT
 Sbjct: 234 TSHEIHVEQKLPQKLSAKAKELPIVDKAPYRFTHGWTYSLNDYFLTRGFASIYVAGVGTR 293

20 Query: 293 GSTGFMTSGDYQQIQSFKAVIDWLNQVTAFTSHKRDQVKANWSNGLVATTGKSYLGTM 352
 S GF TSGDYQQI S AVIDWLNQ+ A+TS K+ ++KA+W+NG VA TGKSYLGTM
 Sbjct: 294 SSDGFQTSQDYQQIYSMTAVIDWLNQRARAYTSRKKTHEIKASWANGKVAMTGKSYLGTM 353

25 Query: 353 STGLATTGVEGLKVIIAEAAISTWYDYRENGLVCSPPGGYPGEDLDVLTETYSRNLLAG 412
 + G ATTGVEGL+VI+AEA IS+WY+YYRENGLV SPGG+PGEDLDVL LTYSRNL
 Sbjct: 354 AYGAATTGVEGLEVLAEAGISSWYNYRENGLVRSPPGGFPGEDLDVLAALTYSRNLDGA 413

30 Query: 413 DYIKNNDCYQALLNEQSKAIDRQSGDYNQYWHDRNYLTHVNNVKS RVVYTHGLQDWNVPK 472
 D++K N Y+ L E + A+DR+SGDYNQ+WHDRNYL + + VK+ V+ HGLQDWNV P
 Sbjct: 414 DFLKGNAEYEKRLAEMTAALDRKSGDYNQFWHDRNYLINTDKVKADVLIVHGLQDWNVTP 473

35 Query: 473 RHVYKVFNALPQTIKKHLFLHQGHVYMHNNQSIDFRESMNALLSQELLGIDNHFQLEEV 532
 Y + ALP+ KH FLH+G H+YM++WQSIDF E++NA +LL D + L V
 Sbjct: 474 EQAYNFWKALPEGHAKHAFHHRGAHIYMNQSWQSIDFSETINAYFVAKLLDRDLNLNLPPV 533

40 Query: 533 IWQDNTTEQWQVLDAGFGGNHQEQIGLGD---SKKLIDNHYDKEAFDTYCKDFNVFKNDL 589
 I Q+N+ +Q W +++ FG N Q ++ LG S DNHYD E F Y KDFNVFK DL
 Sbjct: 534 ILQENSKDQVWMTMMNDFGANTQIKLPLGKTAVSFAQFDNHYDDETFKKYSKDFNVFKDL 593

45 Query: 590 FKGNNKNTQITINLPLKKNYLLNGQCKLHLRVKTSDDKAILSAQILDYGPKKRFDKDTPTI 649
 F+ NK N+ I+L L +NG +L LR+K +D K LSAQILD+G KKR +D +
 Sbjct: 594 FE--NKANEVIDLELPSMLTINGPVELELRLKLNNTKGFLSAQILDGQKKRLEDKARV 651

50 Query: 650 KFLNSLDNGKNFAREALRELPTKDHYRVISKGVLNLQNRDILLTIEAIEPEQWFDIEFS 709
 K LD G+NF + L ELP + Y++I+KG NLQN+ +LLT+ ++ ++WF I+F
 Sbjct: 652 KDFKVLDRGRNFMDDLVELPLVESPYQLITKGFTNLQNG-NLLTVSDLKADEWFTIKFE 710

55 Query: 710 LQPSIYQLSKGDNLRILYTTDFEHTIRDNASYSITVDLSQSYLTIP 756
 LQP+IY L K D LR+ILY+TDFEHT+RDN + +DLSQS L IP
 Sbjct: 711 LQPTIYHLEKADKLRLVILYSTDFEHTVRDNRKVITYEIDLSQSKLIIP 757

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4991> which encodes the amino acid sequence <SEQ ID 4992>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2553(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 481/758 (63%), Positives = 587/758 (76%), Gaps = 4/758 (0%)

Query: 1 MRYNQFSYIPTKPNFAFEELKGLGFPLNKKNSDKANLEAFLRHSFLNQTDYALSLLIV 60
 MRYNQFSYIPT A EELK LGF L+ + + KA+LE+FLR F + D+DY LS LI
 Sbjct: 1 MRYNQFSYIPTSLERAEEELKELGFDLQLQKTAKASLESFLRKLFFHYPDSDYPLSHLIA 60

-1803-

Query: 61 DAKTDALTFKSNSDLTLENLQWIYQLLGFIPFVDFKDPKAFLODINFVSYDN--IFQ 118
 DAL+FF+S +L+ E + LQ+LGFIP VDF + AFL + FP+ +D I +
 Sbjct: 61 KNDMDALSFFQSEQELSKEVFDLLALQVLGFIPGVDFTEADAFDLKLAFFIHFDETEI 120

5 Query: 119 SLHLLACRGKSGNTLIDQLVADGLLHADNHYHFFNGKSLATFNTNQLIREVVYVETSLD 178
 +HLLA R KSG TLID LV+ G+L DN YHFFNGKSLATF+T+QLIREVVYVE LD
 Sbjct: 121 HIHLLATRCCKSGMTLIDDLVSQGLMTMDNDYHFFNGKSLATFDTSQLIREVVYVEAPLD 180

10 Query: 179 TMSSGEHDLVKVNIIRPTTEHTIPTMTASPHYHQGINDPAADQKTYQMEGALAVKQPKHI 238
 T G+ DL+KVNIIRP ++ +PT+MT SPYHQGIN+ A D+K Y+ME L VK+ + I
 Sbjct: 181 TDQDQGLDLIKVNIIRPQSQKPLPTLMTSPYHQGINEVANDKKLYRMEKELVVKRRQI 240

15 Query: 239 QVDTKPFKEEVKHPKSLPISPATESFTHIDSYSLNDYFLSRGFANIYVSGVGTAGSTGFM 298
 V+ + F P KLPI ESF++I+SYSLNDYFL+RGFANIYVSGVGTAGSTGFM
 Sbjct: 241 TVEDRDFIPLETQPCKLPIGQNLESFSYINSYSLNDYFLARGFANIYVSGVGTAGSTGFM 300

20 Query: 299 TSGDYQQIQSFKAVIDWLNKGVTAPTSHKRDQVKANWSNGLVATTGKSYLGMTSTGLAT 358
 TSG+Y QI+SFKAVIDWLN+ TA+TSH + QV+A+W+NLGV TTGKSYLGMTSTGLAT
 Sbjct: 301 TSGNYAQIESFKAVIDWLNGRATAYTSHSKTHQVRADWANGLVCTTGKSYLGMTSTGLAT 360

25 Query: 359 TGVEGLKVIIAEEAISTWYDYRENGLVCSPPGGYPGEDLDVLTETYSRNLLAGDYIKNN 418
 TGV+GL +IIAE+AIS+WY+YYRENGLVCSPPGGYPGEDLDVLTETYSRNLLAGDY+++N
 Sbjct: 361 TGVDGLAMIIAESAISSWYNYRENGLVCSPPGGYPGEDLDVLTETYSRNLLAGDYLRHN 420

30 Query: 419 DCYQALLNEQSKAIDRQSGDYNQYWHDRNYLTHVNNVKSrvvYTHGLQDWNVVKPRHVYKV 478
 D YQ LLN+QS+A+DRQSGDYNQ+WHDRNYL + + +K VVYTHGLQDWNVVKPR VY++
 Sbjct: 421 DRYQELLNQSQALDRQSGDYNQFWHDRNYLKNAHQIKCDVVYTHGLQDWNVVKPRQVYEI 480

35 Query: 479 FNALPQTIKKHLFLHQGHVYMHNVQSIDFRESMNALLSQELLGIDNHQLEEVWQDNT 538
 FNALP TI KHLFLHQG+HVMHNVQSIDFRESMNALL Q+LLG+ N F L E+IWQDNT
 Sbjct: 481 FNALPSTINKHLFLHQGEHVYMHNVQSIDFRESMNALLCQKLLGLANDFSLPEMIWQDNT 540

40 Query: 539 TEQTWQVLDAFGGNHQEQIGLGDSSKLLIDNHYDKEAFDTYCKDFNVFKNDLFKGNKNTNQ 598
 Q WQ FG + +++ LG LIDNHY ++ F Y KDF FK LFKG K NQ
 Sbjct: 541 CPQNWQERKVFGTSTIKELDLGQELLLIDNHYGEDEFKAYGKDFRAFKAAALFKG--KANQ 598

45 Query: 599 ITINLPLKKNYLLNGQCKLHLRVKTSKKAISAQILDYGPKKRFDPTTIKFLNLSLNG 658
 I++ L+++ +NG+ L L+VK+S+ K +LSAQILDYG KKR D P +S+DNG
 Sbjct: 599 ALIDILLEEDLPINGEIVLQLKVKSSSENKGLLSAQILDYGKKRLGDLPIALTQSSIDNG 658

Query: 659 KNFAREALRELPTTKDHYRVISKGVNLQNRDILLTIEAIEPEQWFDIEFSLQPSIYQLS 718
 +NF+RE L+ELPF +D YRVISKG +NLQNR +L +IE I +W + LQP+IY L
 Sbjct: 659 QNFSREPLKELPFREDSYRVISKGFNMNLQNRNLSIETIPNNKWMTVRLPLQPTIYHLE 718

Query: 719 KGDNLRIILYTTDFEHTIRDNASYSITVDLSQSYLTIP 756
 KGD LR+ILYTTDFEHT+RDN++Y++T+DLSQS L +P
 Sbjct: 719 KGDTLRVILYTTDFEHTVRDNSNYALTIDLSQSLIVP 756

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1618

A DNA sequence (GBSx1713) was identified in *S.agalactiae* <SEQ ID 4993> which encodes the amino acid sequence <SEQ ID 4994>. This protein is predicted to be PrfA. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3976(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1804-

A related GBS nucleic acid sequence <SEQ ID 10125> which encodes amino acid sequence <SEQ ID 10126> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:CAA65740 GB:X97014 PrfA [Listeria seeligeri]
    Identities = 54/181 (29%), Positives = 95/181 (51%), Gaps = 1/181 (0%)

    Query: 38 DYTILKDGIVKQSVLSKYGTEFNLRYVTGLEITSILNTDYSQHMGEFYNVRIESETAHF 97
              +Y L +G+ K + +S+ G NL+Y G I D + +G YN+ + SE A
    Sbjct: 36 EYCIFLHEGVAKLTSISESGDILNLQYYKGAFIIMTGFIDTEKSLGY-YNLEVVSEQAAA 94

10  Query: 98 YKVRRTFLKDIINNDIELQGYVKDFYHNRLKESMKMKMCMLTNGRIGAISTQLYDLSKMF 157
              Y ++ S + ++ D++ Y+ D ++ S+ K +NG++G+I Q L+ ++
    Sbjct: 95 YIIKISDLKELVSKDLKQLFYIIDTLQKQVSYSLAKFNDFFSSNGKVGSI CGQFLILAYVY 154

15  Query: 158 GEERDNGDIYIN FVITNEELGKFCGISTGSSVSRLKQLKDDHIIRIEKQHIIITNVEKLGK 218
              GEE NG +T +ELG GI+ S+VSRI+ +LK +++I + + I N+ LK
    Sbjct: 155 GEETPNGIKITLEKLTMOELGCSSGIAHSSAVSRIISKLKQENVIEYKDSYFYIKNIAYLK 215

```

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4995> which encodes the amino acid sequence <SEQ ID 4996>. Analysis of this protein sequence reveals the following:

```

    Possible site: 13
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
25      bacterial cytoplasm --- Certainty=0.4088(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

30  Identities = 186/223 (83%), Positives = 203/223 (90%)

    Query: 1 MEEVMNHQILQNYINSHNLPPIIEKDYHKYLTFFESLEEDYTYILKDGIVKQSVLSKYGTEF 60
              +E+ +NH ILQ YI++HN PII EK YHKYLTFFESLEED+TYILKDGIVKQSVLSKYG EF
    Sbjct: 17 LEKSVNHILQRYIDNHNFPPIIEKSYHKYLTFFESLEEDFTYILKDGIVKQSVLSKYGMEF 76

35  Query: 61 NLRYVTGLEITSILNTDYSQHMGEFYNVRIESETAHFYKVRRTFLKDIINNDIELQGYVK 120
              NLRYVTGLEITS+LNT YS+ MGEFYNVRIESE A FYKVRRS FLKDIN DIELQGYVK
    Sbjct: 77 NLRYVTGLEITSVLNTGYSKDMGEFYNVRIESEKASFYKVRRSAFLKDINEDIELQGYVK 136

40  Query: 121 DFYHNRLKESMKMKMCMLTNGRIGAISTQLYDLSKMFGEERDNGDIYIN FVITNEELGKF 180
              DFYHNRL+KSMKKMCMLTNGRIGAISTQ+YDL +FGEE NG I INFVITNEELGKF
    Sbjct: 137 DFYHNRLQKSMKKMCMLTNGRIGAISTQIYDMLTLFGEELPNGQILINFVITNEELGKF 196

    Query: 181 CGISTGSSVSRLKQLKDDHIIRIEKQHIIITNVEKLGKDHIVF 223
              CGIST SSVSRILKQLK+ +IIRI+KQHIIITN++KLKD+IVF
45  Sbjct: 197 CGISTASSVSRLKQLKEKNIIRIDKQHIIITNLDKLDKNIVF 239

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 1619

A DNA sequence (GBSx1714) was identified in *S.agalactiae* <SEQ ID 4997> which encodes the amino acid sequence <SEQ ID 4998>. Analysis of this protein sequence reveals the following:

```

    Possible site: 46
    >>> Seems to have an uncleavable N-term signal seq
55  INTEGRAL    Likelihood = -14.33    Transmembrane 167 - 183 ( 159 - 193)
    INTEGRAL    Likelihood = -7.96     Transmembrane 18 - 34 ( 10 - 37)
    INTEGRAL    Likelihood = -7.75     Transmembrane 373 - 389 ( 369 - 392)

```

-1805-

```

INTEGRAL    Likelihood = -5.68    Transmembrane  214 - 230 ( 212 - 234)
INTEGRAL    Likelihood = -4.78    Transmembrane  243 - 259 ( 241 - 262)
INTEGRAL    Likelihood = -2.71    Transmembrane   48 -  64 (  47 -  65)
INTEGRAL    Likelihood = -2.60    Transmembrane  283 - 299 ( 283 - 300)

```

```

----- Final Results -----

```

```

      bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB15662 GB:Z99122 similar to antibiotic resistance protein
[Bacillus subtilis]

```

```

Identities = 106/401 (26%), Positives = 199/401 (49%), Gaps = 21/401 (5%)

```

```

Query: 3   DKLFPNKHFIGITILNFIVMVYYLFTVIIAFIATKELGVSTSQAGLATGIYIVGTLIARL 62
          D ++ K FI + ++N V++ +Y F ++      +ELG + SQ GL  +++++ +I R
Sbjct: 5   DAIWTKDFIMVLLVNLFFVFFFTFLTVLPFIYTLQELGGTESQCGLLISLFLLSAIIITRP 64

Query: 63  IFGKQLEVLGRKLVLRGGAIFFYLLTTLAYFYMPISIGMYLVRFINGFGYGVSTATNTIV 122
          G +E G+K +      + L++ Y + + ++ +RF G + +++T T I
Sbjct: 65  FSGAIVERFGKKRMAIVSMALFALSSFLYMPIHNFSLLLGLRFFQGIWFSILTTVTGAIA 124

Query: 123 TAYIPADKRGEGINFYGLSTSLAAAIGPFVGTFMLDNLHINFKMVIVLCSILIAIVVLGA 182
          IPA +RGE+ ++ +S +LA AIGPF+G ++      ++F +      ++ + +L +
Sbjct: 125 ADIIPAKRRGEGLGYFAMSMNLAMAIGPFLGLNLMRV--VSFPVFFTAFAFMVAGLLVS 182

Query: 183 FVFPVKNITLNPEQLAKSKSWTIDSF-----IEKKAIFITIIAFLMGISYASVLGFQKLY 237
          F+ V      +K T+ F      EK A+ I + +      Y++V + ++
Sbjct: 183 FLIKVPQ-----SKDSGTTVPRFAFSDMFEKGALKIATVGLFISFCYSTVTSYLSVF 234

Query: 238 TTEINLMTVGAYFFIVYALVITLTPRSMGRMDAKGDKWVLYPSYFLTLGLLALGSAMG 297
          ++L + YFF+ +A+ + + RP G+L D G V+YPS L ++GL +L
Sbjct: 235 AKSVDLSDISGYFFVCFVATMMIARPFTGKLFDKVGP GIVYIYPSILIFSGLCMLSFTHS 294

Query: 298 SVTYLLSGALIGFGYGTTFMSCGQAASIKGVEEHRFNTAMSTYMIGLDLGLGAGPYILGLV 357
          + LLGA+IG GYG+ + C Q +I+      HR A +T+ D G+ G Y+ GL
Sbjct: 295 GLMLLLSGAVIGLGYGSIVPCMQLAIQKSPAHRSGFATATFFTTFFDSGIAVGSYVFG- 353

Query: 358 KDGFLGAGVQSFRELEFWIAAIIPVVGILYFLKSSRQVETK 398
          F+ +      F ++ A + ++ +LY      + E +
Sbjct: 354 ---FVASA--GFSAIYLTAGLFVLIALLLYTWSQKKPAEAE 389

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4999> which encodes the amino acid sequence <SEQ ID 5000>. Analysis of this protein sequence reveals the following:

```

Possible site: 35

```

```

>>> Seems to have an uncleavable N-term signal seq

```

```

INTEGRAL    Likelihood =-12.31    Transmembrane  202 - 218 ( 194 - 225)
INTEGRAL    Likelihood = -7.80    Transmembrane   53 -  69 (  44 -  71)
INTEGRAL    Likelihood = -7.17    Transmembrane  407 - 423 ( 404 - 426)
INTEGRAL    Likelihood = -5.26    Transmembrane  249 - 265 ( 247 - 269)
INTEGRAL    Likelihood = -3.77    Transmembrane  279 - 295 ( 276 - 297)
INTEGRAL    Likelihood = -2.23    Transmembrane   11 -  27 (  10 -  27)
INTEGRAL    Likelihood = -2.13    Transmembrane   83 -  99 (  82 -  99)
INTEGRAL    Likelihood = -1.91    Transmembrane  312 - 328 ( 311 - 328)

```

```

----- Final Results -----

```

```

      bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:CAB15662 GB:Z99122 similar to antibiotic resistance protein
[Bacillus subtilis]

```

-1806-

Identities = 110/390 (28%), Positives = 194/390 (49%), Gaps = 11/390 (2%)

Query: 38 EKLEFNKHFVAITVINFIVMVYLLFTVIIAFVATRELGAQTSQAGLATGIYILGTLRL 97
 + ++ K F+ + ++N V++ +Y F ++ +ELG SQ GL +++L ++ R
 Sbjet: 5 DAIWTKDFIMVLLVNLVVFVFFFTFLTVLPITYTLQELGGTESQGGLLISLFLLSAIIITRP 64

Query: 98 IFGKQLEVFGRRLVLRGGAIFYLLTTLAYFYMPITISMMYLVRFLNGFGYGVVSTATNTIV 157
 G +E FG++ + + L++ Y + S++ +RF G + +++T T I
 Sbjet: 65 FSGAIVERFSGKKRMAIVSMALFALSSFLYMPIHNFSLLLGLRPFQGIWFSILTTVTGAIA 124

Query: 158 TAYIPARKRGEINFYGLSTSLAAAIGPFVGTFLMDNLHIDFRMIIVLCSVLIGCVVGA 217
 IPA++RGE+ ++ +S +LA AIGPF+G ++ + F + ++ + ++ +
 Sbjet: 125 ADIIPAKRRGEGLYGFAMSMNLAMAIGPFLGLNLMRV--VSFPVFFTAFAFMVAGLLVS 182

Query: 218 FAFPVKNMSLNAEQLAQTKSWTVDSFIEKKALFITAI AFLMGIAYASVLGFQKLYTSEIH 277
 F V + + + + EK AL I + + Y++V + ++ +
 Sbjet: 183 FLIKVPQSKDSGTTVFR---FAFSDMFEKGALKIATVGLFISFCYSTVTSYLSVFAKSVD 239

Query: 278 LTTVGAYFFVYALIIITITRPAMGRMDAKGDKWVLYPSYFLAMGLFLLGSVSSGGSYL 337
 L+ + YFFV +A+ + I RP G+L D G V+YPS L ++GL +L SG L
 Sbjet: 240 LSDISGYFFVCFVATMMIARPFQKGLFDKVGPGIVITYPSILIFSGLCMLSFTHSGMLLL 299

Query: 338 LSGALIGFGYGTFSMCSQAASIQGVDEHRFNTAMSTYMIGLDLGLGAGPYLLGLIKDLAL 397
 LSGA+IG GYG+ + C Q +IQ HR A +T+ D G+ G Y+ GL
 Sbjet: 300 LSGAVIGLGYGSIVPCMQLAIQKSPAHRSGFATATFTFFDSDGIAGSVYVGLF----- 354

Query: 398 GSGVASFRHLFWLAAVIPLICTLLYLLKTK 427
 A F ++ A + LI LLY K
 Sbjet: 355 -VASAGFSAIYLTAGLFVLIALLLYTWSQK 383

An alignment of the GAS and GBS proteins is shown below.

Identities = 328/396 (82%), Positives = 370/396 (92%), Gaps = 1/396 (0%)

Query: 1 MEDKLEFNKHFIFIGITILNFIVMVYLLFTVIIAFIATKELGVSTSQAGLATGIYIVGTLIA 60
 ME+KLEFNKHF+ IT++NFIVMVYLLFTVIIAF+AT+ELG TSQAGLATGIYI+GTL+A
 Sbjet: 36 MEEKLEFNKHFVAITVINFIVMVYLLFTVIIAFVATRELGAQTSQAGLATGIYILGTLIA 95

Query: 61 RLIFGKQLEVLGRKLVLRGGAIFYLLTTLAYFYMPISIGVMYLVRFLNGFGYGVVSTATNT 120
 RLIFGKQLEV GR+LVLRGGAIFYLLTTLAYFYMP+I +MYLVRFLNGFGYGVVSTATNT
 Sbjet: 96 RLIFGKQLEVFGRRLVLRGGAIFYLLTTLAYFYMPITISMMYLVRFLNGFGYGVVSTATNT 155

Query: 121 IVTAYIPADKRGEINFYGLSTSLAAAIGPFVGTFLMDNLHINFKMVIVLCSILIAIVVL 180
 IVTAYIPA KRGEINFYGLSTSLAAAIGPFVGTFLMDNLHI+F+M+IVLCS+LI VV+
 Sbjet: 156 IVTAYIPARKRGEINFYGLSTSLAAAIGPFVGTFLMDNLHIDFRMIIVLCSVLIGCVV 215

Query: 181 GAFVFPVKKNITLNPEQLAKSKSWTIDSFIEKKAIFITIIAFLMGISYASVLGFQKLYTTE 240
 GAF FPKN++LN EQLAK+KSWT+DSFIEKKA+FIT IAFLMGI+YASVLGFQKLYT+E
 Sbjet: 216 GAFAPVKNMSLNAEQLAQTKSWTVDSFIEKKALFITAI AFLMGIAYASVLGFQKLYTSE 275

Query: 241 INLMTVGAYFFIVYALVITLTPRSMGRMDAKGDKWVLYPSYFLTLGLALLGSAMGSVT 300
 I+L TVGAYFF+VYAL+IT+TRP+MGRMDAKGDKWVLYPSYFL +GL LLGS +
 Sbjet: 276 IHLTTVGAYFFVYALIIITITRPAMGRMDAKGDKWVLYPSYFLAMGLFLLGSVSSGGS 335

Query: 301 YLLSGALIGFGYGTFSMCSQAASIKGVEHRFNTAMSTYMIGLDLGLGAGPYLLGLVKDG 360
 YLLSGALIGFGYGTFSMCSQAASI+GV+EHFRFNTAMSTYMIGLDLGLGAGPY+LGL+KD
 Sbjet: 336 YLLSGALIGFGYGTFSMCSQAASIQGVDEHRFNTAMSTYMIGLDLGLGAGPYLLGLIKDL 395

Query: 361 FLGAGVQSFRELFWIAAIIPVVCGLYFLKS-SRQV 395
 LG+GV SFR LFW+AA+IP++C +LY LK+ +RQV
 Sbjet: 396 ALGSGVASFRHLFWLAAVIPLICTLLYLLKTKTRQV 431

A related GBS gene <SEQ ID 8863> and protein <SEQ ID 8864> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: 8.26

-1808-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1620

- 5 A DNA sequence (GBSx1715) was identified in *S.agalactiae* <SEQ ID 5001> which encodes the amino acid sequence <SEQ ID 5002>. Analysis of this protein sequence reveals the following:

Possible site: 46
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0151(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06903 GB:AP001518 unknown conserved protein [Bacillus halodurans]
Identities = 52/143 (36%), Positives = 84/143 (58%)

20 Query: 5 YERILIAIDGSYESELAVEKGINVALRND AELL LTHVIDAHAYQSEGVS DYVFDRQEQE 64
Y IL+A+DGS +++ A+ K N A A+L + HVID+ ++ + + V E +
Sbjct: 2 YNHILVAVDGSTQAKRALYKAFNYAKEFKADLFICHVIDSRSFATVEQYDRTVVGA AELD 61

25 Query: 65 SADVLAYFEKLAHSGKGLTKIKKITEIGNPKTLLAKDIPREKADLIMVGATGLNTFERLL 124
+L + + A G+ K+ I + G+PK ++K I + DLI+ GATGLN ER L
Sbjct: 62 GKLLQRYSEEA EKAGVDKVHTILDFGSPKANISKTTIAQKYDIDLITGATGLNAVERFL 121

30 Query: 125 IGSTSEYILRH SKVDMLVVRDSK 147
+GS SE + RH+K D+L+VR+ +
Sbjct: 122 MGSVSES VARHAKCDVLIVRNDQ 144

There is also homology to SEQ ID 3658:

Identities = 105/150 (70%), Positives = 121/150 (80%)

35 Query: 1 MTQKYERILIAIDGSYESELA VEKGINVALRND AELL LTHVIDAHAYQSEGVS DYVFDR 60
M+ KY+RIL+AIDGSYESELA KG+NVALRND A LLL HVID A QS F Y++++
Sbjct: 31 MSLKYKRILVAIDGSYESELA FNKGVNVALRNDATLLLVHVIDTRALQSVATFD TYIYEK 90

40 Query: 61 QE QESADVLAYFEKLAHSGKGLTKIKKITEIGNPKTLLAKDIPREKADLIMVGATGLNTF 120
EQE+ DVL FEK A G+T IK+I E GNPK LLA DIP RE ADLIMVGATGLNTF
Sbjct: 91 LEQEAQDVLD DFEKQAQIAGITNIKQIIEFGNPKNLLAHDIPDRENADLIMVGATGLNTF 150

45 Query: 121 ERL LIGSTSEYILRH SKVDMLVVRDSK KTL 150
ERLLIGS+SEYI+RH+K+D+LVVRDS KTL
Sbjct: 151 ERL LIGSSSEYIMRHAKIDLLVVRDSTKTL 180

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1621

- 50 A DNA sequence (GBSx1716) was identified in *S.agalactiae* <SEQ ID 5003> which encodes the amino acid sequence <SEQ ID 5004>. This protein is predicted to be glycerol uptake facilitator protein (glpF). Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -8.65 Transmembrane 261 - 277 (257 - 281)
55 INTEGRAL Likelihood = -5.73 Transmembrane 201 - 217 (199 - 222)

-1809-

```

INTEGRAL    Likelihood = -4.51    Transmembrane  92 - 108 (  91 - 110)
INTEGRAL    Likelihood = -4.30    Transmembrane  44 -  60 (  42 -  62)
INTEGRAL    Likelihood = -2.18    Transmembrane  15 -  31 (  11 -  31)
INTEGRAL    Likelihood = -1.54    Transmembrane 150 - 166 ( 149 - 166)

```

----- Final Results -----

```

bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA25231 GB:M58315 putative [Lactococcus lactis]
Identities = 183/290 (63%), Positives = 228/290 (78%), Gaps = 10/290 (3%)

```

```

Query: 2  IEITWTVKYITEFIATAFLIILGNGAVANVDLKGTKGNNSGWIIIAIGYGLGVMPALMF 61
+++TWTVKYITEF+ TA LII+GNGAVANV+LKGTK + W+II GYGLGVM+PA+ F
Sbjct: 1  MDVTWTVKYITEFVGTALLIIMGNGAVANVELKGTKAHAQSWMIIGWGYGLGVMLPAVAF 60

```

```

Query: 62  GNVSGNHINPAFTLGLAFSGLFPPWAHVQYILAQILGAMFGQLVVMVYQPYFVKTENPN 121
GN++ + INPAFTLGLA SGLFPPWAHV QYI+AQ+LGAMFGQL++VMVY+PY++KT+NPN
Sbjct: 61  GNIT-SQINPAFTLGLAASGLFPPWAHVAQYIIAQVLGAMFGQLLIVMVYRPYYLKTQNP 119

```

```

Query: 122 HVLGSFSTISALDDGQKSSRKAAYINGFLNEFVGSFVLFPGALALTKNYFGVE----LVG 177
+LG+FSTI +DD + +R A INGFLNEF+GSFVLFPGA+A T +FG + +
Sbjct: 120 AILGTFSTIDNVDDNSEKTRLGATINGFLNEFLGSFVLFPGAVAATNIFGSGSITWMTN 179

```

```

Query: 178 KLVQAGYDQTTAATRISPYVTGSLA-----VAHLGIGFLVMTLVASLGPGTGPALNPARD 232
L G D +++ +V S A +AHL +GFLVM LV +LGGPTGP LNPARD
Sbjct: 180 YLKGQGADVSSSDVMNQIWWQASGASASKMIAHLFLGFLVMGLVVALGGPTGPGLNPA 239

```

```

Query: 233 LGPRIVHRLLEPKQILGQAKEDSKWYAWVPVLAPIVASILAVALFKLLYL 282
GPR+VH LLEPK +LG+AK SKWYAWVPVLAPI+AS+ AVALFK++YL
Sbjct: 240 FGPRLVHSLLEPKSVLGEAKGSSKWYAWVPVLAPILASLAVALFKMIYL 289

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5005> which encodes the amino acid sequence <SEQ ID 5006>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have an uncleavable N-term signal seq

```

INTEGRAL    Likelihood = -9.18    Transmembrane 293 - 309 ( 288 - 314)
INTEGRAL    Likelihood = -7.43    Transmembrane  2 -  18 (  1 -  20)
INTEGRAL    Likelihood = -7.38    Transmembrane 233 - 249 ( 228 - 256)
INTEGRAL    Likelihood = -5.57    Transmembrane 124 - 140 ( 123 - 142)
INTEGRAL    Likelihood = -2.87    Transmembrane  76 -  92 (  75 -  93)
INTEGRAL    Likelihood = -2.18    Transmembrane  47 -  63 (  43 -  63)
INTEGRAL    Likelihood = -1.54    Transmembrane 182 - 198 ( 181 - 198)

```

----- Final Results -----

```

bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAA25231 GB:M58315 putative [Lactococcus lactis]
Identities = 176/290 (60%), Positives = 228/290 (77%), Gaps = 10/290 (3%)

```

```

Query: 34  MEMTWTVKYITEFIATAFLIILGNGAVANVDLKGTKGHNSGWLVIAGYGLGVMPALMF 93
M++TWTVKYITEF+ TA LII+GNGAVANV+LKGTK H W++I +GYGLGVM+PA+ F
Sbjct: 1  MDVTWTVKYITEFVGTALLIIMGNGAVANVELKGTKAHAQSWMIIGWGYGLGVMLPAVAF 60

```

```

Query: 94  GNVSGNHINPAFTVGLAVSGLFPPWAHVLYVVAQLLGAIFGQLVVMVYKPYFMKTENPN 153
GN++ + INPAFT+GLA SGLFPPWAHV QY++AQ+LGA+FGQL++VMVY+PY++KT+NPN
Sbjct: 61  GNIT-SQINPAFTLGLAASGLFPPWAHVAQYIIAQVLGAMFGQLLIVMVYRPYYLKTQNP 119

```

```

Query: 154 HVLGSFSTISSLDNGQKDSHKASYINGFLNEFVGSFVLFPGALALTKNYFGVELVGKLE 213

```

-1810-

```

      +LG+FSTI ++D+  + +  +  INGFLNEF+GSFVLFFGA+A T  +FG + + +
Sbjct: 120 AILGTFSTIDNVDDNSEKTRLGATINGFLNEFLGSFVLFFGAVAATNIFFGSQSITWMTN 179

Query: 214 -----AGYDQTTAATQISPYVTGSLA---VAHIGIGFLVMVLVTSLGGPTGPPALNPARD 264
      A      +      QI      +G+ A      +AH+ +GFLVM LV +LGGPTGP LNPARD
Sbjct: 180 YLKGQGADVSSSDVMNQIWWQASGASASKMIAHLFLGFLVMGLVVALGGPTGPPALNPARD 239

Query: 265 FGPRLLHHFLPKSVLGQAKGDSKWWYAWVPVAPILAAIVAVAAFKYLYI 314
      FGPRL+H  LPKSVLG+AKG SKWWYAWVPV+APILA++ AVA FK +Y+
Sbjct: 240 FGPRLVHSLLPKSVLGAEKSSKWWYAWVPVLAPILASLAVALFKMIYL 289

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 240/281 (85%), Positives = 267/281 (94%)

```

Query: 2  IEITWTVKYITEFIATAFLIILGNGAVANVDLKGTKGNNSGWIIIAIGYGLGVMMPALMF 61
      +E+TWTVKYITEFIATAFLIILGNGAVANVDLKGTKG+NSGW++IA GYGLGVMMPALMF
Sbjct: 34  MEMTWTVKYITEFIATAFLIILGNGAVANVDLKGTKGHNSGWLVIAFGYGLGVMMPALMF 93

Query: 62  GNVSGNHINPAFTLGLAFSGLFPWAHVQYILAQILGAMFGQLVVMVYQPYFVKTENPN 121
      GNVSGNHINPAFT+GLA SGLFPWAHV QY++AQ+LGA+FGQLVVMVY+PYF+KTENPN
Sbjct: 94  GNVSGNHINPAFTVGLAVSGLFPWAHVLYVVAQLLGAIFGQLVVMVYKPYFMKTENPN 153

Query: 122 HVLGSFSTISALDDGQKSSRKAAYINGFLNEFVGSFVLFFGALALTKNYFGVELVGKLVQ 181
      HVLGSFSTIS+LD+GQK S KA+YINGFLNEFVGSFVLFFGALALTKNYFGVELVGKL++
Sbjct: 154 HVLGSFSTISSLDNGQKDSHKASYINGFLNEFVGSFVLFFGALALTKNYFGVELVGKLVQ 213

Query: 182 AGYDQTTAATRISPYVTGSLAVAHLGIGFLVMTLVASLGGPTGPPALNPARDLGPRIHVRL 241
      AGYDQTTAAT+ISPYVTGSLAVAH+GIGFLVM LV SLGGPTGPPALNPARD GPR++H
Sbjct: 214 AGYDQTTAATQISPYVTGSLAVAHIGIGFLVMVLVTSLGGPTGPPALNPARDFGPRLLHHF 273

Query: 242 LPKQILGQAKEDSKWWYAWVPVLAPIVASILAVALFKLLYL 282
      LPK +LGQAK DSKWWYAWVPV+API+A+I+AVA FK LY+
Sbjct: 274 LPKSVLGQAKGDSKWWYAWVPVAPILAAIVAVAAFKYLYI 314

```

A related GBS gene <SEQ ID 8865> and protein <SEQ ID 8866> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 8
McG: Discrim Score:      2.81
GvH: Signal Score (-7.5): -3.6

```

```

Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
ALOM program count: 6 value: -8.65 threshold: 0.0

```

```

INTEGRAL Likelihood = -8.65 Transmembrane 261 - 277 ( 257 - 281)
INTEGRAL Likelihood = -5.73 Transmembrane 201 - 217 ( 199 - 222)
INTEGRAL Likelihood = -4.51 Transmembrane 92 - 108 ( 91 - 110)
INTEGRAL Likelihood = -4.30 Transmembrane 44 - 60 ( 42 - 62)
INTEGRAL Likelihood = -2.18 Transmembrane 15 - 31 ( 11 - 31)
INTEGRAL Likelihood = -1.54 Transmembrane 150 - 166 ( 149 - 166)
PERIPHERAL Likelihood = 2.92 72

```

modified ALOM score: 2.23

*** Reasoning Step: 3

----- Final Results -----

```

bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

ORF01006(304 - 1146 of 1446)
EGAD|14239|14211(1 - 289 of 289) hypothetical 30.9 kd protein in pepx 5'region {Lactococcus
lactis} SP|P22094|YDP1_LACLC HYPOTHETICAL 30.9 KDA PROTEIN IN PEPX 5'REGION (ORF1).
GP|455286|gb|AAA25206.1|M35865 ORF1 (put.); putative {Lactococcus lactis}
GP|149527|gb|AAA25231.1|M58315 putative {Lactococcus lactis} PIR|B43747|B43747

```

[illegible]

Example 1622

Possible site: 44

```
>>> Seems to have an uncleavable N-term signal seq
```

----- Final Results -----

The protein has no significant homology with any sequences in the GENPEPT database.

A related sequence was also identified in GAS <SEQ ID 9177> which encodes the amino acid sequence
60 <SEQ ID 9178>. Analysis of this protein sequence reveals the following:

-1812-

Possible cleavage site: 21
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -10.77	Transmembrane	139 - 155 (133 - 161)
INTEGRAL	Likelihood = -8.28	Transmembrane	245 - 261 (240 - 269)
INTEGRAL	Likelihood = -7.48	Transmembrane	269 - 285 (263 - 289)
INTEGRAL	Likelihood = -7.06	Transmembrane	97 - 113 (83 - 125)
INTEGRAL	Likelihood = -6.10	Transmembrane	173 - 189 (169 - 194)
INTEGRAL	Likelihood = -1.44	Transmembrane	200 - 216 (200 - 217)

----- Final Results -----
 bacterial membrane --- Certainty=0.531(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 225/301 (74%), Positives = 263/301 (86%)

Query: 10 LTVSLFFCRDLIMNETLLHLGIQLILIIAMIITFYQIVRHRSQKINPFKRFFTGWLWIGF 69
 LT +FFC+L MNE L+L IQ +L+ AM+ F+ +V+H++ KINPFKR+TG WIG
 Sbjct: 1 LTAKVFFCKLVFMNEMLLRLIQLALLVSAMLFIFFMVLVKHLKKNKINPFKRFTGFWIGL 60

Query: 70 VTDALDTLGIGSFATTTTFFKLTKLVEDDRKIPATMTAAHVLPVLLQSLCFIFVVKVEAL 129
 +TDALDTLGIGSFATTTT FKLTKLV DDR++P TMT AHVLPVL+QSLCFIFVVKVE L
 Sbjct: 61 LTDALDTLGIGSFATTTTCFKLTKLVTDRLPGTMTVAHVLPVLIQSLCFIFVVKVEVL 120

Query: 130 TLITMAGAAFIGAFVGA KM TKNWHAPTQVQRILGTLITAAIIMLYRMITNPGAGISDSVH 189
 TL+ MA AAFIGA+ G +TKNWHAPTQVQRILG+LLI AAIIM+ R+I +PG +SD++H
 Sbjct: 121 TLLAMAAAAFIGAYFGTHITKNWHAPTQVQRILGSLIIIAIIMIIRIYHPGEHLSDTIH 180

Query: 190 GLHGIWLVFGIGFNFIIGVLTMTGLGNYAPELIFFSLMGLSPAVAMPVMMMLDAMIMTAS 249
 GLHGIWLVFGIGFNFI+GVLMTMTGLGNYAPELIFFSLMGLSP VAMPVMMMLDAMIMTAS
 Sbjct: 181 GLHGIWLVFGIGFNFIIVGVLMTMTGLGNYAPELIFFSLMGLSPTVAMPVMMMLDAMIMTAS 240

Query: 250 STQFIKSGRVNWNMGFAGLVGTGGILGVIVAVLFTNLDLNSLKTLLVVGIVLFTGAMLIRSSF 310
 S+QFIK+ RV+W+GFAG+V+GGI+GV++AV FLTNLD+NSLK LV+ IV FTG MLIRSSF
 Sbjct: 241 SSQFIKANRVSWDGFAGIVSGGIIGVLLAVFFLTNLDLNSLKLLVIAIVFFTGGMILIRSSF 301

A related GBS gene <SEQ ID 8867> and protein <SEQ ID 8868> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: 2.32
 GvH: Signal Score (-7.5): -5.59
 Possible site: 44
 >>> Seems to have an uncleavable N-term signal seq

ALOM program count: 8 value: -8.70 threshold: 0.0

INTEGRAL	Likelihood = -8.70	Transmembrane	266 - 282 (262 - 290)
INTEGRAL	Likelihood = -7.96	Transmembrane	25 - 41 (24 - 50)
INTEGRAL	Likelihood = -6.42	Transmembrane	110 - 126 (105 - 140)
INTEGRAL	Likelihood = -6.26	Transmembrane	194 - 210 (190 - 215)
INTEGRAL	Likelihood = -5.47	Transmembrane	290 - 306 (289 - 310)
INTEGRAL	Likelihood = -4.35	Transmembrane	128 - 144 (127 - 147)
INTEGRAL	Likelihood = -3.29	Transmembrane	157 - 173 (156 - 174)
INTEGRAL	Likelihood = -2.76	Transmembrane	221 - 237 (221 - 240)
PERIPHERAL	Likelihood = 3.87		67

modified ALOM score: 2.24

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1813-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5009> which encodes amino acid sequence <SEQ ID 5010>:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

```

5      INTEGRAL    Likelihood = -10.77    Transmembrane  151 - 167 ( 145 - 173)
      INTEGRAL    Likelihood = -9.13     Transmembrane   22 - 38 ( 15 - 42)
      INTEGRAL    Likelihood = -8.28     Transmembrane  257 - 273 ( 252 - 281)
      INTEGRAL    Likelihood = -7.48     Transmembrane  281 - 297 ( 275 - 301)
      INTEGRAL    Likelihood = -7.06     Transmembrane  109 - 125 ( 95 - 137)
10     INTEGRAL    Likelihood = -6.10     Transmembrane  185 - 201 ( 181 - 206)
      INTEGRAL    Likelihood = -1.44     Transmembrane  212 - 228 ( 212 - 229)
      INTEGRAL    Likelihood = -0.27     Transmembrane   5 - 21 ( 5 - 21)

```

----- Final Results -----

```

15     bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS sequences follows:

```

20     Score = 405 bits (1029), Expect = e-115
      Identities = 198/301 (65%), Positives = 228/301 (74%)

Query: 1   LTAKVFFCKLVFMNEMILIRLIQALLVSAMLFIFFMVLVKHLKKNKINPFKRFTWTFWIGL 60
      LT +FFC+L MNE L+L IQ +L+ AM+ F+ +V+H++ KINPFKRFT+TG WIG
25     Sbjct: 10 LTVSLFFCRLDIMNETLLHLGILILIIAMIITFYQIVRHRSQKINPFKRFTGLWIGF 69

Query: 61  LTDALDTLGIGSFATTTTCFKLTKLVTDRLPGTMTVAHVLPVLIQSLCFIFVVKVEVX 120
      +TDALDTLGIGSFATTTT FKLTKLV DDR++P TMT AHVLPVL+QSLCFIFVVKVE
30     Sbjct: 70 VTDALDTLGIGSFATTTTFFKLTKLVEDDRKIPATMTAAHVLPVLLQSLCFIFVVKVEAL 129

Query: 121 XXXXXXXXXXXXFIGAYFGTHITKNWHAPTQVQRIILGSLXXXXXXXXXXXXXHPGEHLSDTIH 180
      FIGA+ G +TKNWHAPTQVQRIILG+LL +PG +SD++H
35     Sbjct: 130 TLITMAGAAFIGAFVGAKMTKNWHAPTQVQRIILGTLITAAIIMLYRMITNPGAGISDSVH 189

Query: 181 GLHGIWLFVVGIGFNFI+GVLMTMGLGNYAPELIFFSLMGLSPTVAMPVMMMLDAMIMTAS 240
      GLHGIWLFVVGIGFNFI+GVLMTMGLGNYAPELIFFSLMGLSP VAMPVMMMLDAMIMTAS
40     Sbjct: 190 GLHGIWLFVVGIGFNFI+GVLMTMGLGNYAPELIFFSLMGLSPAVAMPVMMMLDAMIMTAS 249

Query: 241 SSQFIKANRVSWDXXXXXXXXXXXXXXXXXXFSLTNLDINSLKLLVIAIVFTTGGMLIRSSF 301
      S+QFIK+ RV+W+ FLTNLD+NSLK LV+ IV FTG MLIRSSF
45     Sbjct: 250 STQFIKSGRVNWNNGFAGLVTGGILGVIVAVLFLTNLDINSLKTLVVGIVLFTGAMLIRSSF 310

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 1623

A DNA sequence (GBSx1718) was identified in *S.galactiae* <SEQ ID 5011> which encodes the amino acid sequence <SEQ ID 5012>. This protein is predicted to be C3-degrading proteinase. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

50     bacterial cytoplasm --- Certainty=0.2851(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD37110 GB:AF112358 C3-degrading proteinase [Streptococcus pneumoniae]

-1814-

Identities = 92/240 (38%), Positives = 142/240 (58%), Gaps = 11/240 (4%)

Query: 12 PVLRVNNDLNIAFYQESLGFKLISEENAIIVFSAWQNEASFIIEESPTYRTRAVNGTK 71
 P L+ NNR LN FY E+IG K + EE+A E ++EE+P+ RTR V G K
 5 Sbjct: 11 PTLKANNRKLNETFYIETLGMKALLEESAFLSLGDTGTGLE-KLVLEEAPSMRTRKVEGRK 69

Query: 72 KLAKIIVKSQDAKDIEKLLANGAQAIQVYQGQNGYAYETVSPEGDLFLLHAEDDLSQLVA 131
 KLA++IVK ++ +IE +L+ ++Y+GQNGYA+E SPE DL L+HAEDD++ LV
 10 Sbjct: 70 KLARLIVKVENPLEIEGILSKTDSIHRLYKQNGYAFEIFSPEDDLILIHAEDDIASLVE 129

Query: 132 I-ERPELEKKDDTTGLSNFAFQSSISLNVDAVKAEAFYDKVFAGKFPINLSFKEAQQGQDL 190
 + E+PE + + LS F S+ L++P + E+F + + + +L F AQQGQDL
 15 Sbjct: 130 VGEKPEFQTDLASISLSKFEI-SMELHLPTDI--ESFLE---SSEIGASLDFIPAQQGQDL 183

Query: 191 QIAPNETWDIEILECCVNEDTNLNDLKSTFESLGLDVYLDSEKILVISDTSNIEIWISK 250
 + TWD+ +L+ VNE ++ L+ FES + ++ EK + D +N+E+W +
 15 Sbjct: 184 TVDNTVTWDLMLKFLVNE-LDIASLRQKFES--TEYFIPKSEKFFLGKDRNNVELWFEE 240

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5013> which encodes the amino acid
 20 sequence <SEQ ID 5014>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3267(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 130/250 (52%), Positives = 177/250 (70%)

Query: 1 MTLFHSLSLTFKHPVLRVNNRDLNIAFYQESLGFKLISEENAIIVFSAWQNEASFIIEESP 60
 MTL ++TFK PVL RVN+RDLNIAFYQ +LG +L+SEENAIA+FS+W + F+IEESP
 35 Sbjct: 1 MTLMENITFKTPVLRVNDRLNIAFYQNNLGLRLVSEENAIAIFSSWGEGQECFVIEESP 60

Query: 61 TYRTRAVNGTKKLAKIIVKSQDAKDIEKLLANGAQAIQVYQGQNGYAYETVSPEGDLFLL 120
 + RTRAV G KK+ I++K+ K+IE+LLA+GA +++GQNGYA+ET+SPEGD FLL
 Sbjct: 61 SVRTRAVEGPKKVNTIVIKTNQPKIEIQLLAHGAHYDALFKGQNGYAFETISPEGDRFLL 120

Query: 121 HAEDDLSQLVAIERPELEKKDDTTGLSNFAFQSSISLNVDAVKAEAFYDKVFAGKFPINL 180
 HAE D+ L + P LEK GL+ F F I LNV +++AFY +F+ + PI +
 40 Sbjct: 121 HAEQDIKHLQGTDLPSLEKDATFKGLTQFKFDIIVLNVISEERSKAFYRDLFSDQLPITM 180

Query: 181 SFKEAQQGQDLQIAPNETWDIEILECCVNEDTNLNDLKSTFESLGLDVYLDSEKILVISD 240
 F + +G DL I P+ WD+EILE V++D ++ LK+T E G VY+D K K+LV+SD
 45 Sbjct: 181 DFIQEEGPDLAIDPHIAWDLIELEFQVSKDYDMKVLKATLEEDGHKVYIDKKHKVLVLS 240

Query: 241 TSNIEIWISK 250
 S IE+W +K
 50 Sbjct: 241 PSQIEVWFTK 250

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1624

A DNA sequence (GBSx1719) was identified in *S.agalactiae* <SEQ ID 5015> which encodes the amino
 acid sequence <SEQ ID 5016>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1815-

bacterial cytoplasm --- Certainty=0.2510(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAC16441 GB:AL450165 putative esterase [Streptomyces coelicolor]
 Identities = 89/323 (27%), Positives = 143/323 (43%), Gaps = 51/323 (15%)

Query: 10 NTVLELIKEQIKDNLVHGASLAIY-ENGWEHEHYLGT-----IDGNEKVKAGLVYDLA 61
 +T+ EL+ E + + GA+ ++ G + GT +DG++ V+DLA
 Sbjct: 2 STLAELLAEGREQRICSGAAWSVGGPQGPGLDRGWTGTRCWDGPPLDGDD-----VWDLA 55

Query: 62 SVSKVVGVTLLAKLVYQGTIDIDKPLRYYPYTFH--HQTITVRQLATHSSGIDPFIP- 117
 SV+K + G ++ LV +G + +D + Y P + LTVRQL H+SGI +P
 Sbjct: 56 SVTKPIA-GLVVMALVERGALGLDDTVGGYLPDYRGGDKAELTVRQLLAHTSGIPGQVPL 114

Query: 118 NRDQLNATQLKDAINHIKVLDEKSFK--YTDINFLLLGFMLEEVLGDSLDKLFKRYIFTP 175
 RD L +A+ + + + Y+ F++LG + E G+ L+ L +R + P
 Sbjct: 115 YRDHPTRAALLEAVRLLEPLTAQPGTRVQYSSQGFIVLGLIAEAAAGEPLEALVERLVCA 174

Query: 176 FQMKETSFGPRVEAVPTVVGIND-----GIVHDPKAKVLGKHTGSAGLFSTIDDLQ 226
 +++T F P V D G VHD A VLG G AGLFST+ D++
 Sbjct: 175 LGLRDTVFRPDAGRARRAVATEDCPWRGRRVVGVEVDENAVVLGGVGGHAGLFSTLADME 234

Query: 227 RFSIHYL-----KDDFA-KPLWNNYSLSKSRSLAWD-----IDKDWINHT 265
 R + FA + L+ R+LAW + HT
 Sbjct: 235 RLGAALAAGGRGLLRPETFALMTAAHTDGLALRRALAWQGRDPVGSPPAGEVFGPESYGH 294

Query: 266 GYTGFPIALNYQKQAAAIFLTNR 288
 G+TG + ++ + A+ LTNR
 Sbjct: 295 GFTGTSLWVDPATRRYAVLLTNR 317

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3885> which encodes the amino acid sequence <SEQ ID 3886>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.33 Transmembrane 57 - 73 (57 - 74)

----- Final Results -----
 bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 174/302 (57%), Positives = 229/302 (75%), Gaps = 1/302 (0%)

Query: 9 TNTVLELIKEQIKDNLVHGASLAIYENGWEHEHYLGTIDGNEKVKAGLVYDLASVSKVVG 68
 T V++ I+ + +Y GASLA++++G W E+++GTIDG V A LVYDLASVSKVVG
 Sbjct: 6 TLAVIKCIENHLHKVYKGASLALFQSGRWQYHIGTIDGRRPVDANLVYDLASVSKVVG 65

Query: 69 VGTLLAKLVYQGTIDIDKPLRYYPYTFHHQTITVRQLATHSSGIDPFIPNRDQLNATQLK 128
 V T+ L+ GT+ +D PL+ YYP+ T+T+RQL TH+SG+DP+IPNRD LNA QL+
 Sbjct: 66 VATICNILLNNGTLALDDPLKVYPSIADATVTTIRQLLTHTSGLDPIIPNRDVINAQQQLR 125

Query: 129 DAINHIKVLDEKSFKYTDINFLLLGFMLEEVLGDSLDKLFKRYIFTPFQMKETSFGPRVE 188
 A+NH+ E+K+F YTD+NFLLLGFMLEE+ +SLD++F + IFTPF M TSFGPR E
 Sbjct: 126 KALNHLTQKENKNFYTDVNFLLLGFMLEELFSESLDQIFDKTIFTPFGMYHTSFGPRPE 185

Query: 189 AVPTVVGINDGIVHDPKAKVLGKHTGSAGLFSTIDDLQRFISIHYLKDDFAKPLWNNYSL 248
 AVPT+ G++DG VHDPAK+L KH+GSAGLFST+ DL+ FS HYL D F+ LW NYS
 Sbjct: 186 AVPTLKGVSDEVDHPKAKILKKHSGSAGLFSTLADLESFNNHYLNDPFSCLWRNYSQQ 245

Query: 249 K-SRSLAWDIDKDWINHTGYTGPFIALNYQKQAAAIFLTNRFTSYDDRPLWIKRRHVQE 307
 RSL W++D DWI+HTGYTGP+ LN ++Q AAIFLTNR+ DD+ W+K+R+ +

-1816-

Sbjct: 246 TIERSLGWNLDGDWISHTGYTGPFLMLNKKEQTAAIFLTNRITYDEDDKSKWLKERQLLYN 305

Query: 308 AI 309

A+

5 Sbjct: 306 AL 307

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1625

10 A DNA sequence (GBSx1720) was identified in *S.agalactiae* <SEQ ID 5017> which encodes the amino acid sequence <SEQ ID 5018>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0935(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA25177 GB:D21804 FMN-binding protein [Desulfovibrio vulgaris]

Identities = 53/124 (42%), Positives = 76/124 (60%), Gaps = 2/124 (1%)

25 Query: 1 MLNKKFLQVLKYEGVVSITSWIELAPHVTNTWNSYL/TITDDQRILAPAAGMTHLENDLNN 60

ML F +VLK EGVV+I + E PH+ NTWNSYL + D RI+ P GM E ++

Sbjct: 1 MLPGTFFFEVLKNEGVVAIATQGEDGPHLVNTWNSYLVLDGNRIVVPVGGMHKTEANVAR 60

Query: 61 NSKIIMTLGSRVEGRDGYQGTGFRIEGTAKLLEAGSDFEIVKEKYPFLRKVLEVTPIINV 120

+ +++MTLGSR+V GR+G GTGF I G+A G +FE + ++ + R L +T ++

30 Sbjct: 61 DERVLMITLGSRKVAGRNG-PGTGFLIRGSAAFRTDGPEFEAI-ARFKWARAALVITVWSA 118

Query: 121 IQLL 124

Q L

Sbjct: 119 EQTL 122

35

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1626

40 A DNA sequence (GBSx1721) was identified in *S.agalactiae* <SEQ ID 5019> which encodes the amino acid sequence <SEQ ID 5020>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3799(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-1817-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1627

A DNA sequence (GBSx1722) was identified in *S.agalactiae* <SEQ ID 5021> which encodes the amino acid sequence <SEQ ID 5022>. Analysis of this protein sequence reveals the following:

Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3175(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10123> which encodes amino acid sequence <SEQ ID 10124> was also identified.

The protein has homology to a pyruvate formate-lyase from *S.mutans*:

>GP:BAA09085 GB:D50491 Pyruvate formate-lyase [Streptococcus mutans]
Identities = 709/770 (92%), Positives = 750/770 (97%)

20 Query: 7 MATVKTINTDIFEQAWEGFKGVDWKEKASIAIRFVQANYAPYDGDSEFLAGATERSLHIKKV 66
MATVKTINTD+FE+AWEGFKG DWK++ASI+RFVQ NY PYDG ESFLAG TERSLHIKKV
Sbjct: 1 MATVKTINTDVFEKAWEGFKGTDWKDRASISRFDQNTPTPYDGGESFLAGPTERSLHIKKV 60

25 Query: 67 IEETKAHYEETRFPMMDTRVASISELPAGFIDKDNELIFGIQNDLFLKLNFMKPGGIRMAE 126
+EETKAHYEETRFPMMDTR+ SI+++PAG+IDK+NELIFGIQNDLFLKLNFMKPGGIRMAE
Sbjct: 61 VEETKAHYEETRFPMMDTRITSIADIPAGYIDKENELIFGIQNDLFLKLNFMKPGGIRMAE 120

30 Query: 127 TTLKENGVEPDPAVHEIFTKYATTVNDGIFRAYTSNIRRARHAHTVTGLPDAYSRGRIIG 186
T LKE+GYEPDPAVHEIFTKYATTVNDGIFRAYTSNIRRARHAHTVTGLPDAYSRGRIIG
Sbjct: 121 TALKEHGYEPDPAVHEIFTKYATTVNDGIFRAYTSNIRRARHAHTVTGLPDAYSRGRIIG 180

35 Query: 187 VYARLAVYGADYLMQEKVNDWNALNDIDEESIRLREEINLQYQALGEVVKLGDLYGVDVR 246
VYARLA+YGADYLMQEKVNDWN++ +IDEESIRLREEINLQYQALGEVV+LGDLYG+DVR
Sbjct: 181 VYARLALYGADYLMQEKVNDWNSIAEIDEESIRLREEINLQYQALGEVVRLGDLYGLDVR 240

40 Query: 247 KPAMNTKEAIQWVNIAMAVCRVINGAATSLGRVPIVLDFIAERDLARGTFTSEIQEFV 306
KPAMN KEAIQW+NIAAMAVCRVINGAATSLGRVPIVLDFIAERDLARGTFTSEIQEFV
Sbjct: 241 KPAMNVKEAIQWINIAMAVCRVINGAATSLGRVPIVLDFIAERDLARGTFTSEIQEFV 300

45 Query: 307 DDFVLKLRITVKFARTKAYDALYSGDPTFITTS MAGMGADGRHRVTKMDYRFLNTLDNIGN 366
DDFV+KLRTVKFARTKAYD LYSGDPTFITTS MAGMGADGRHRVTKMDYRFLNTLDNIGN
Sbjct: 301 DDFVMKLRITVKFARTKAYDELYSGDPTFITTS MAGMGADGRHRVTKMDYRFLNTLDNIGN 360

50 Query: 367 SPEPNLTVLWSDQLPYAFRRYCMSMSHKHSSIQYEGVSTMAKEGYGEMSCISCCVSPIDP 426
+PEPNLTVLWS +LPY+FR YCMSMSHKHSSIQYEGV+TMAKEGYGEMSCISCCVSPIDP
Sbjct: 361 APEPNLTVLWSSKLPYSFRHYCMSMSHKHSSIQYEGVTTMAKEGYGEMSCISCCVSPIDP 420

55 Query: 427 ENEDKRHNLQYFGARVNVKALLTGLNGGYDDVHKDYKVFDDIPRDEVLNFDTVKANFE 486
ENED+RHNLQYFGARVNV+KALLTGLNGGYDDVHKDYKVFDD+PIRDEVL+F+TVKANFE
Sbjct: 421 ENEDRRHNLQYFGARVNVKALLTGLNGGYDDVHKDYKVFDDVEPIRDEVLDFETVKANFE 480

60 Query: 487 KSLDWLTDITYVDAMNIIHYMTDKYNYEAVQMAFLPSHVVRANMGFGICGFANTVDSLSAIK 546
K+LDWLTDITYVDAMNIIHYMTDKYNYEAVQMAFLP+ V+ANMGFGICGF+NTVDSLSAIK
Sbjct: 481 KALDWLTDITYVDAMNIIHYMTDKYNYEAVQMAFLPTRVKANMGFGICGFSNTVDSLSAIK 540

Query: 547 YATVKPIRDEDDGYIYDYETVGDFFRYGEDDDRVDLSIAEWLLEAFHGR LAKHKLYKDAEAT 606
YATVKPIRDEDDGYIYDYETVG+FFRYGEDDDRVDLSIAEWLLEAFH RLA+HKLYKDAEAT
Sbjct: 541 YATVKPIRDEDDGYIYDYETVGNFFRYGEDDDRVDLSIAEWLLEAFHTRLARHKLYKDEAT 600

Query: 607 VSLLTITSNVAYSQQTGNSPVHKGVYLNEDGSVNLSKVEFFSPGANPSNKAKGGWLQNLN 666

-1818-

VSLLTITSNVAYSQKTGNSPVHKGVYLNEDGSVNLSKVEFFSPGANPSNKA GGWLQNLN
 Sbjct: 601 VSLLTITSNVAYSQKTGNSPVHKGVYLNEDGSVNLSKVEFFSPGANPSNKASGGWLQNLN 660

Query: 667 SLKLDFAHANDGISLTTQVSPRALGKTFDEQVDNLVTVLDGYFENGQGHVNLNVMDLKD 726
 5 SL KLDFAHANDGISLTTQVSP+ALGKTFDEQV NLVT+LDGYFE GGQHVNLNVMDLKD
 Sbjct: 661 SLKKLDFAHANDGISLTTQVSPKALGKTFDEQVANLVTTILDGYFEGGQGHVNLNVMDLKD 720

Query: 727 VYDKIMNGEDVIVRISGYCVNTKYLTPEQKTELQTVFHEVLSMDALTN 776
 VYDKIMNGEDVIVRISGYCVNTKYLT EQKTELQTVFHEVLSMDA T+
 10 Sbjct: 721 VYDKIMNGEDVIVRISGYCVNTKYLTKEQKTELQTVFHEVLSMDAATD 770

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5023> which encodes the amino acid sequence <SEQ ID 5024>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3184(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 701/773 (90%), Positives = 742/773 (95%), Gaps = 1/773 (0%)

Query: 2 FKEKTMATVKTNITDIFEQAWEGFGVDWKEKASIRFVQANYAPYDGDESFLAGATERSL 61
 FKEK MATVKTNITD+FE+AWEGFGK DWKEKAS++RFVQANY PYDGDESFLAGATERSL
 Sbjct: 5 FKEKFMATVKTNITDVFEKAWEGFGKTDWKEKASVSRLFVQANYTPYDGDESFLAGATERSL 64

Query: 62 HIKKVIETKAHYEETRFPMOTRVASISELPAGFIDKDNELIFGIQNDLFLNFMFKGG 121
 HIKKVIETKAHYE TRFP DTR SI+++PAGFIDK+NELI+GIQNDLFLNFMFKGG
 30 Sbjct: 65 HIKKVIETKAHYEATRFYPDTRPTSADIAPAGFIDKENELIYGIQNDLFLNFMFKGG 124

Query: 122 IRMAETTLKENGYPDPAPVHEIFTKYATTVNDGIFRAYTSNIRRARAHTVTGLPDAYS 181
 IRMAETTLKENGYPDPAPVHEIFTKY TTVNDGIFRAYTSNIRRARAHTVTGLPDAYS
 35 Sbjct: 125 IRMAETTLKENGYPDPAPVHEIFTKYVTTVNDGIFRAYTSNIRRARAHTVTGLPDAYS 184

Query: 182 GRIIGVYARLAVYGADYLMQEKVNDWNAIDDEESIRLREEINLQYQALGEVVKLGDL 241
 GRIIGVYARLA+YGADYLMQEKVNDWNA+ +IDDEESIRLREE+NLQYQALGEVVKLGDL
 40 Sbjct: 185 GRIIGVYARLALYGADYLMQEKVNDWNAITEIDDEESIRLREEVNLQYQALGEVVKLGDL 244

Query: 242 GVDVRKPAMNTKEAIQWVNIAMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTSE 301
 GVDVR+PA N KEAIQWVNIAMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTSE
 Sbjct: 245 GVDVRRPAQNVKEAIQWVNIAMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTSE 304

Query: 302 IQEFVDDFVLKLRVTKFARTKAYDALYSGDPTFITTSMAGMGADGRHRVTKMDYRFLNTL 361
 IQEFVDDFVLKLRVTKF RTKAYDALYSGDPTFITTSMAGMG DGRHRVTKMDYRFLNTL
 45 Sbjct: 305 IQEFVDDFVLKLRVTKFRTKAYDALYSGDPTFITTSMAGMGNDGRHRVTKMDYRFLNTL 364

Query: 362 DNIGNSPEPNLTVLWSDQLPYAFRRYCMMSHKHSSIQYEGVSTMAKEGYGEMSCISCCV 421
 DNIGNSPEPNLTVLW+DQLP FRRYCM MSHKHSSIQYEGV+TMAKEGYGEMSCISCCV
 50 Sbjct: 365 DNIGNSPEPNLTVLWTDQLPETFRYCMMSHKHSSIQYEGVTTMAKEGYGEMSCISCCV 424

Query: 422 SPLDPENEDKRHNLQYFGARVNVKALLTGLNGGYDDVHKDYKVF- IDPIRDEVLNFD 480
 SPLDPENE++RHN+QYFGARVNV+KALLTGLNGGYDDVH+DYKVF+ ++PI EVL +D
 55 Sbjct: 425 SPLDPENEEQRHNIQYFGARVNVKALLTGLNGGYDDVHRDYKVFNVVEPITSEVLEYDE 484

Query: 481 VKANFEKSLDWLTDYVDAMNIIHYMTDKYNYEAVQMAFLPSHVRANMGFGICGFANTVD 540
 V ANFEKSLDWLTDYVDA+NIIHYMTDKYNYEAVQMAFLP+H RANMGFGICGFANTVD
 60 Sbjct: 485 VMANFEKSLDWLTDYVDALNIIHYMTDKYNYEAVQMAFLPTHQRANMGFGICGFANTVD 544

Query: 541 LSIAIKYATVKPIRDEGDIYDYETVGDFFRYGEDDDRVDSIAEWLLEAFHGRILAKHKLY 600
 +LSIAIKYATVK IRDE+GYIYDYE GDFPRYGEDDDRV D IA+WL+EA+H RLA HKLY
 Sbjct: 545 TLSIAIKYATVKTIREDENGYIYDYETVGDFFRYGEDDDRVDDIAKWLMEAYHTRLASHKLY 604

Query: 601 KDAEATVSLTITSNVAYSQKTGNSPVHKGVYLNEDGSVNLSKVEFFSPGANPSNKA 660

-1819-

K+AEA+VSLLTITTSNVAYSKQTGNSPVH+GV+LNEDG+VN S+VEFFSPGANPSNKA KGG
 Sbjct: 605 KNAEASVSLLTITTSNVAYSKQTGNSPVHRGVFLNEDGTVNTSQVEFFSPGANPSNKA KGG 664

Query: 661 WLQNLNSLSKLDFAHANDGISLTTQVSPRALGKTFDEQVDNLVTVLDGYFENGQGHVNLN 720
 WLQNLNSL+KL+F+HANDGISLTTQVSPRALGKTFDEQVDNLVTVLDGYFENGQGHVNLN
 Sbjct: 665 WLQNLNSLAKLEFSHANDGISLTTQVSPRALGKTFDEQVDNLVTVLDGYFENGQGHVNLN 724

Query: 721 VMDLKDVDYDKIMNGEDVIVRISGYCVNTKYLTPKTELTQRVFHEVLSMDDA 773
 VMDL DVYDKIMNGEDVIVRISGYCVNTKYLTPKTELTQRVFHEVLSMDDA
 Sbjct: 725 VMDLNDVDYDKIMNGEDVIVRISGYCVNTKYLTPKTELTQRVFHEVLSMDDA 777

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1628

A DNA sequence (GBSx1723) was identified in *S. galactiae* <SEQ ID 5025> which encodes the amino acid sequence <SEQ ID 5026>. This protein is predicted to be DNA-damage inducible protein P (dinP). Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10121> which encodes amino acid sequence <SEQ ID 10122> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF95431 GB:AE004300 DNA-damage-inducible protein P [Vibrio cholerae]
 Identities = 136/349 (38%), Positives = 210/349 (59%), Gaps = 14/349 (4%)

Query: 12 INDTSRKIIHIDMDAFFASVEERDNPSLKGPVIIGSDPRKTGGRGVVSTCNYEARKFGV 71
 + D RKIIH+DMD FFA+VE RDNP+ + + +G ++ RGV+STCNY+ARKFGV
 Sbjct: 1 MQDRIRKIIHVDMDCFFAAVEMRDNPAYREIALAVGGHEKQ---RGVISTCNYQARKFGV 57

Query: 72 HSAMSSKEAYERCPQAFISGNYQKYRQVGMVERDIFKKYTDLVEPMSIDEAYLDVTENK 131
 SAM + +A + CPQ + G Y+ V +++ IF++YT L+EP+S+DEAYLDV+E+
 Sbjct: 58 RSAMPTAQLKLCPLHVVPGRMVSYSVSQQIQTIFQRYTSLIEPLSLDEAYLDVSEST 117

Query: 132 MGIKSAVKLAKMIQYDIWNDVHLTCSAGISYNKFLAKLASDFEKPGLTLILPDQAQDFL 191
 SA +A+ I+ DIW +++LT SAG++ KFLAK+ASD KP GL ++ PD+ Q+ +
 Sbjct: 118 AYQGSATLIAQAIRRDWQELNLTASAGVAPIKFLAKVASDLNKPGLYVVTPDKVQEMV 177

Query: 192 KPLPIEKFHGVGKRSVEKLHALGVYTGEDLLSLSEISLIDMFGFRFGYDLYRKARGINASP 251
 LP+EK GVGK ++EKLH G+Y G D+ L+ FGR G L++K+ GI+
 Sbjct: 178 DSLPLEKIPGVGKVALEKLHQAGLYVGADVRRADYRKLLHQFGRGLGASLWKKSHGIDERE 237

Query: 252 VKPDRVRKSGISEKTYGKLLYNEADIKAEISKNVQRRVVASLEKNKKVGKTIV---LKVRY 308
 V +R RKS+G E T+ + + + I + + + + + I+ +KV++
 Sbjct: 238 VVTERERKSVGVEYTFQSNISTFQECWQVIEQKLYPELDARLSRAHPQRGIIKQGIKVKF 297

Query: 309 ADFETLTTRMTLEEYTDQF--QIIDQVAKAIFDTEESVFGIRLLGVTV 355
 ADF+ T D+ ++++QV + IRLLG++V
 Sbjct: 298 ADFQQTTEHVHPALELDYFHELLEQV-----LTRQQGREIRLLGLSV 340

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5027> which encodes the amino acid sequence <SEQ ID 5028>. Analysis of this protein sequence reveals the following:

Possible site: 27

-1820-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1921(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 276/363 (76%), Positives = 323/363 (88%)

10 Query: 6 MLIFPLINDTSRKIIHIDMDAFFASVEERDNP SLKGKPVIIIGSDPRKTGGRGVVSTCNYE 65
 MLIFPLINDTSRKIIHIDMDAFFA+VEERDNP+LKKGKPV+IG DPR+TGGRGVVSTCNYE
 Sbjct: 1 MLIFPLINDTSRKIIHIDMDAFFAAVEERDNPALKKGKPVVIGKDPRETGGRGVVSTCNYE 60

15 Query: 66 ARKFGVHSAMSSKEAYERCPQAIFISGNYQKYRQVGMEVRDIFKKYTDLVEPMSIDEAYL 125
 ARK+G+HSAMSSKEAYERCP+AIFISGNY+KYR VG ++R IFK+YTD+VEPMSIDEAYL
 Sbjct: 61 ARKYGHSAMSSKEAYERCPKAIFISGNYEKYRTVGDQIRIRIFKRYTDVVEPMSIDEAYL 120

20 Query: 126 DVTENKMGISAVKLAQMIQYDIWNDVHLTCSAGISYNKFLAKLASDFEKPGLTLILPD 185
 DVT+NK+GKISAVK+AK+IQ+DIW +V LTCSAG+SYNKFLAKLASDFEKP GLTL+L +,
 Sbjct: 121 DVTDNKLGKISAVKIAKLIQHDIWKEVGLTCSAGVSYNKFLAKLASDFEKPGLTLVLKE 180

25 Query: 186 QAQDFLKLPIEKFHGVGKRSVEKLHALGVYTGEDLLSLSEISLIDMFGFRFGYDLYRKAR 245
 A FL LPIEKFHGVGK+SV+KLH +G+YTG+DLL++ E++LID FGRFG+DLYRKAR
 Sbjct: 181 DALCFLAKLPIEKFHGVGKKS VKKLHDMGIYTGQDLLAVPEMTLIDHFGFRFGDLYRKAR 240

30 Query: 246 GINASPVKPDVRKRSIGSEKTYGKLLYNADIAEISKNVQRVVASLEKNKKVGKTIVLK 305
 GI+ SPVK DR+RKSIGSE+TY KLLY E DIKAEISKNV+RV A L+ +KK+GKTIVLK
 Sbjct: 241 GISNSPVKYDRIRKSIGSERTYAKLLYQETDIKAEISKNVKRVAAALQDHKKLGKTIVLK 300

35 Query: 306 VRYADFTLTTRMTLEEYTDQFQIIDQVAKAIFDTLEESVFGIRLLGVTVTTLNEHEAI 365
 VRYADF TLTKR+TL E T++ I+QVA IFD+L E+ GIRLLGVT+T LE++ I
 Sbjct: 301 VRYADFTTLTKRVTLPETRNAAQIEQVAGDIFDSLSENPAIRLLGVTMTNLEDKVADI 360

Query: 366 YLD 368
 LD
 Sbjct: 361 SLD 363

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1629

A DNA sequence (GBSx1724) was identified in *S.agalactiae* <SEQ ID 5029> which encodes the amino acid sequence <SEQ ID 5030>. Analysis of this protein sequence reveals the following:

Possible site: 41

45 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -13.11 Transmembrane 70 - 86 (58 - 92)
 INTEGRAL Likelihood = -5.20 Transmembrane 105 - 121 (100 - 123)
 INTEGRAL Likelihood = -4.25 Transmembrane 126 - 142 (123 - 144)
 50 INTEGRAL Likelihood = -2.71 Transmembrane 18 - 34 (18 - 34)

----- Final Results -----

55 bacterial membrane --- Certainty=0.6243(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5031> which encodes the amino acid sequence <SEQ ID 5032>. Analysis of this protein sequence reveals the following:

Possible site: 32

-1821-

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -13.00 Transmembrane 69 - 85 (62 - 93)
 INTEGRAL Likelihood = -6.85 Transmembrane 16 - 32 (11 - 37)
 INTEGRAL Likelihood = -4.30 Transmembrane 99 - 115 (96 - 121)
 INTEGRAL Likelihood = -3.66 Transmembrane 126 - 142 (121 - 143)

----- Final Results -----

bacterial membrane --- Certainty=0.6201(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 57/155 (36%), Positives = 96/155 (61%), Gaps = 5/155 (3%)

Query: 1 MVSYEKVRRLRTATITIIIVLSLVLFTGISVQLAKTEI-NKGNTGNLPKEHIEAV 59
 M+SYEKVR++L+T+TI II+LN L +V L + +++++ N+ L E + +
 Sbjct: 1 MISYEKVRQALKTSTIAIIILNGLGVLSLMGFAGIFYLQSQLKNEAFRAQLTTEQLAQL 60

Query: 60 LSATTPFMLFVTALIVLVNIAIVIFCIKNLRAIKRNQTVNYLPYYLGFAITVGLVILGFL 119
 S+ TPFM+F++ L VL IAI++FC +NL +K+ TV+Y+PY LG ++V ++ F
 Sbjct: 61 QSSMTPFMIFLSVLNVLAIIAIVFCAQNLSKLGKGLTVSYIPYILGLILSVIGLVNQFT 120

Query: 120 TTKAPWAIAINIVFQAIFGLLYFHAYQKAQKLNER 154
 TT + + ++ A++G A+ KA+ LNE+
 Sbjct: 121 TTSMVGTILILIQAALYGF----AFYKAKTLNEK 151

SEQ ID 5030 (GBS227) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 119 (lane 5; MW 21.2kDa).

GBS227-His was purified as shown in Figure 227, lane 8-9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1630

A DNA sequence (GBSx1725) was identified in *S.agalactiae* <SEQ ID 5033> which encodes the amino acid sequence <SEQ ID 5034>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1224(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14706 GB:Z99118 similar to conjugation transfer protein
 [Bacillus subtilis]
 Identities = 328/754 (43%), Positives = 484/754 (63%), Gaps = 25/754 (3%)

Query: 2 EVFFTGTIERIIFENASNFFKILLLEIEDTSDFDDEVIITGTMADVTEGEEYTFWGT 61
 E + GT+ +I+ N +N + +L +++ +T +D V +TG + E E YTF+G +
 Sbjct: 13 EPYLKGTVNTVIYHNDTNLYTVLKVKTETSEAIEDKAVSVTGYPALQEEETYPYFGKI 72

Query: 62 TQHPKYGEQLQSVRYERAKPTSG-GLVKYFSSEQFKGIGKTAQRIVELYGDNTIDKILE 120
 HPK+G Q Q+ +++ PT+ G+++Y SS+ F+GIGKTA+ IV+ GD+ I+KIL
 Sbjct: 73 VTHPKFGLQFQAEHFKEIPTTKEGIIQYLSSDLFEGIGKTAEEIVKKLGDSAINKILA 132

-1822-

Query: 121 SPEQLSTISGLSKINREAFIAKLKLNYGTEQVLAKLAEYGLSNRAAIQIFDHYKEESLEV 180
 L + LSK + L+ + G EQ++ L ++G + +++I+ Y+ E+LE
 Sbjct: 133 DASVLYDVPRLSKKKADTLAGALQRHQGLEQIMISLNQFGFGPQLSMKIYQAYESETLEK 192

5 Query: 181 INENPYQLVEDIQIGIGFKIADQLAEQVGIESDSPKRFRAAIHITLVESSMEQGDITYIEAR 240
 I ENPYQLV+D++GIGF AD+L ++G+ + P+R +AAI++TL + + +G TYIE
 Sbjct: 193 IQENPYQLVKDVEGIGFGKADELGSRMGLSGNHPERVKAAILYLETTCLESGHTYIETE 252

10 Query: 241 TLLEKTTITLEEA-----RQIELDPS---IVAKELTNLIAEDKVQHIGTKIFSNTLFFAE 292
 L+ T +LL ++ R E+D + I E +++ ED + + +LF+AE
 Sbjct: 253 QLIIDTQSLNQSAREGQRITEMDAANAIIALGENKDIVIEDG-----RCYFPSLFYAE 306

15 Query: 293 EGIKKNLQRIILNQP-LDKQLNHKIDIDREIRDIQKSLNIHYDNIQEKAIREALLSKVFIIT 351
 + + K ++ I +Q + Q + + +++ + Y Q++AI++AL S + +LT
 Sbjct: 307 QNVAKRKVHIASQTEYENQFPSEFLLALGELEERMDVQYAPSQKEAIQKALSSPMLLLT 366

20 Query: 352 GPGGTGKTTVINGIIEAYSELHHIDLN---KND--IPIVLAAPTGRAARRMNELTGLPS 405
 GPGGTGKTTVI GI+E Y ELH + L+ K D PIVLAAPTGRAA+RM+E TGLP+
 Sbjct: 367 GPGGTGKTTVIRGIVELYGELHGVSLDPSAYKKDEAFPIVLAAPTGRAAKRMSESTGLPA 426

25 Query: 406 ATIHRHLGLNGSDYQSLDDY-LDCSLIIDEFSMVDTWLANQLFDALDSHTQVIVGDS 464
 TIHR LG NG + +D ++ L+IIDE SM+D WLAN LF A+ H Q+IIVGD
 Sbjct: 427 VTIHRLLGWNGAEGFTHTEDQPIEGKLLIIDEASMLDIWLANHLFKAIPDHIQIIIVGDE 486

30 Query: 465 DQLPSVGPQVLADLLNINALPHVKLEKIFRQSEESTIVTLANQMRQGFLEPEDFTAKKAD 524
 DQLPSVGPQVL DLL +P V+L I+RQ+E S+IV LA+QM+ G LP + TA D
 Sbjct: 487 DQLPSVGPQVLRDLLASQVIPTVRLTDIYRQAEQSSIVELAHQMKNGLLPNNLTAPT KD 546

35 Query: 525 RSYFEASANIIPNMISKIVQSALKSGIEAHEIQILAPMYRGQAGINNLLIMQNLNPLK 584
 RS+ + I ++ K+V +ALK G A +IQ+LAPMYRG+AGIN LN+++Q++LNP K
 Sbjct: 547 RSFIRCGSQIKEVVEKVVANALKKGYTAKDIQVLAPMYRGKAGINELNVMQLDILNPPK 606

40 Query: 585 D-NNQFTFNDINFRIGDKVLHLVNDTELNVFNGDIGYITDLIPAKYTESKQDEIYMTFDG 643
 + + F D+ +R GDK+L LVN E NVFNGDIG IT + AK K+D ++FDG
 Sbjct: 607 EKRRELKFGDVVYRTGDKILQLVNQPENNVFNGDIGEITSIFYAKENTEKEDMAVVSFDG 666

45 Query: 644 QEVIYQRKEWLKITLAYAMSIHKSQGSEFQVILPITRQSGRMLQRNLIYTAITRSKSKL 703
 E+ + +K++ + T AY SIHKSQGSEF +V+LP+ + RML+RNL+YTAITR+K L
 Sbjct: 667 NEMTFTKKDFNQFTHAYCCSIHKSQGSEFFIVVLPVVKGYRMLRRNLLYTAITRAKKFL 726

Query: 704 ILLGEIGAFDFAVKNEGAK-RNTYLIERFENKQE 736
 IL GE A ++ VKN A R T L R + E
 Sbjct: 727 ILCGEEEALEWGVKNNDATVRQTSKLNRLSVQVE 760

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5035> which encodes the amino acid sequence <SEQ ID 5036>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 232-234

The protein has homology with the following sequences in the databases:

>GP:CAB14706 GB:Z99118 similar to conjugation transfer protein
 [Bacillus subtilis]

Identities = 318/769 (41%), Positives = 473/769 (61%), Gaps = 29/769 (3%)

Query: 7 GTVDRIIFENQANFFKILLIAIEDTSDIDDFEIIITGTMAIIIEGDDYTFWGETQHPK 66
 GTV+ +I+ N N + +L + + +T I+D + +TG + E + YTF+G++ HPK
 Sbjct: 18 GTVNTVIYHNDTNLYTVLKVKVTETSEAIEDKAVSVTGYFPALQEETTYTFYGVKIVTHPK 77

-1823-

Query: 67 YGQQLKLSRYQKIKPSSS-GLVNYFSSDHFKGIGKKTAEKIIALYGHNTIDHILEDPSKL 125
+G Q + ++K P++ G++ Y SSD F+GIGKKTAE+I+ G + I+. IL D S L
Sbjct: 78 FGLQFQAEHFKEIPTTKEGIIQYLSSDLFEGIGKKTAEIIVKLGDSAINKILADASVL 137

Query: 126 ETISGLSKANRQAFVAKLKLNYGTEQLIAGLVELGLSNRFALQAFEKYKEEALDLVKENP 185
+ LSK L+ + G EQ++ L + G + +++ ++ Y+ E L+ ++ENP
Sbjct: 138 YDVPRLSKKKADTLAGALQRHQGLEQIMISLNQFGFGPOLSMKIYQAYESETLEKIQENP 197

Query: 186 YQLVEDLQGFQFKMADALAENLGIESDSPKRFRAALLHCLLEESINRGDTYVQARQLLDF 245
YQLV+D++G GF AD L +G+ + P+R +AA+L+ L ++ G TY++ QL+
Sbjct: 198 YQLVKDVEGIGFGKADELGSRMGLSGNHPERVKAAILYTTLETTCLSEGHTYIETEQLIID 257

Query: 246 AITLL-----EDARQVECDPAVAEQLSE---LIIEGKIKNSDTKLFDAISLYFAEEGIAN 297
+LL E R E D A L E ++IE D + + SL++AE+ +A
Sbjct: 258 TQSLLNQSAREGQRITEMDAANAIILALGENKDIVIE-----DGRCYFPSLFYAEQNVAK 311

Query: 298 NISRLLD-TPLSQSFSDHTIQTTIQAVQKDFAITYDQVQOEAITKALTSKVFLLTGGPGT 356
+ + T F + +++ + Y Q+EAI KAL+S + LLTGGPGT
Sbjct: 312 RVKHIASTQTEYENQFPPESEFLLALGELEERMDVQYAPSQKEAIQKALSSPMLLLTGGPGT 371

Query: 357 GKTTVIRGILQAYANLHQIDLD---KKD--LPILLAAPTGRAARMNELTGLPSATIHR 410
GKTTVIRGI++ Y LH + LD KKD PI+LAAPTGRAA+RM+E TGLP+ TIHR
Sbjct: 372 GKTTVIRGIVELYGELHGVSLDPSAYKKDEAFPIVLAAPTGRAAKRMSESTGLPAVTIHR 431

Query: 411 HGLNGDNDYQAMEDY-LDCDLLIVDEFMSVDTWLANQLLGAINSTTQVIVGDSQQLPS 469
LG NG + ED ++ LLI+DE SM+D WLAN L AI Q+IIVGD DQLPS
Sbjct: 432 LLGWNGAEGFTHTEDQPIEGKLLIIDEASMLDIWLANHLFKAIPDHIQIIIVGDEDQLPS 491

Query: 470 VGPGQVLSDLLKVNSLPQIALQKIFRQSQUESTIVNLADQMRRGILAADFRDKKADRSYFE 529
VGPGQVL DLL +P + L I+RQ++ S+IV LA QM+ G+L + DRS+
Sbjct: 492 VGPGQVLRDLLASQVIPTVRLTDIYRQAEQSSIVELAHQMKNGLLPNNLTAPT KD RS FIR 551

Query: 530 AQAAPFIDPMIQKIVLSAISKSGIPAEIQLAPMYKGQAGINHLNQLMQELLN-PLQGQTE 588
+ I ++++K+V +A+K G A++IQ+LAPMY+G+AGIN LN ++Q++LN P + + E
Sbjct: 552 CGGSQIKEVVEKVVANALKKGYTAKDIQVLAPMYRGKAGINELNVMQLDILNPPKEKRRE 611

Query: 589 FLFNDTHFRKGDKVLHLVNDALQNVFNNGDIGYITDLIPAKYTESKQDELILDFDGESEVTY 648
F D +R GDK+L LVN + NVFNNGDIG IT + AK K+D ++ FDG+E+T+
Sbjct: 612 LKFGDVVYRTGDKILQLVNQPENNVFNNGDIGEITSIFYAKENTEKEDMAVVSFDGNEMTF 671

Query: 649 PRNEWLKLTLAYAMSIHKSQSEFQVVLPIRQSGRLLQRNVIYTAITRSKSKLILGE 708
+ ++ + T AY SIHKSQSEF +V+LP+ + R+L+RN++YTAITR+K LIL GE
Sbjct: 672 TKKDFNQFTHAYCCSIHKSQSEFPVIVLPPVKGYYRMLRRNLLYTAITRAKKFLILCGE 731

Query: 709 YTAFEYAIK-HEGDKRQTYLIERFQEQSDLASSQPNQELKSKEQTSLSFS 756
A E+ +K ++ RQT L R Q + + + EL++ ++ FS
Sbjct: 732 EEALEWGVKNNDATVRQTSKLNRLSVQVE----EMDAELEALQKELPFS 776

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 544/816 (66%), Positives = 665/816 (80%), Gaps = 10/816 (1%)

Query: 1 MEVFFTGTIERIIFENASNFFKILLLEIEDTSDSDFDDVEVIITGTMDADVIEGEEYTFWGT 60
ME FTGT++RIIFEN +NFFKILL IEDTSD DD E+IITGTMD+IEG++YTFWG
Sbjct: 1 MEVFTGTVDRIIFENQANFFKILLLAIEDTSDIDDFEIIITGTMDLIEGDDYTFWGE 60

Query: 61 LTQHPKYGEQLQSVRYERAKPTSGGLVKYFSSEQFKGIGKKTARIVELYGDNTIDKILE 120
LTQHPKYG+QL+ RY++ KP+S GLV YFSS+ FKGIGKKTAA++I+ LYG NTID ILE
Sbjct: 61 LTQHPKYGQQLKLSRYQKIKPSSSGLVNYFSSDHFKGIGKKTAEKIIALYGHNTIDHILE 120

Query: 121 SPEQLSTISGLSKINREAFIAKLKLNYGTEQVLAKLAEYGLSNRAAIQIFDHYKEESLEV 180
P +L TISGLSK NR+AF+AKLKLNYGTEQ++A L E GLSNR A+Q F+ YKEE+L++
Sbjct: 121 DPSKLETISGLSKANRQAFVAKLKLNYGTEQLIAGLVELGLSNRFALQAFEKYKEEALDL 180

Query: 181 INENPYQLVEDIQGIGFKIADQLAEQVGIESDSPKRFRAAIIHTLVESSEMQGDTYIEAR 240
+ ENPYQLVED+QG GFK+AD LAE +GIESDSPKRFRAA++H L+E S+ +GDTY++AR
Sbjct: 181 VKENPYQLVEDLQGFQFKMADALAENLGIESDSPKRFRAALLHCLLEESINRGDTYVQAR 240

-1824-

5 Query: 241 TLLEKTTITLLEEARQIELDPSTIVAKELTNLIAEDKVQHIGTKIFSENTLFFAEEGIKKNLQ 300
 LL+ ITLLE+ARQ+E DP+ VA++L+ LI E K+++ TK+F +L+FAEEGI N+
 Sbjct: 241 QLLDFAITLLEDARQVECDPAVAEQLSELIIEGKIKNSDTKLFDAISLYFAEEGIANNIS 300

10 Query: 301 RILNQPLDKQLNKHKIDIREIRDIQKSLNIHYDNIQEKAIREALLSKVFIITGGPGTGKTT 360
 R+L+ PL + +H I I+ +QK I YD +Q++AI +AL SKVF+ITGGPGTGKTT
 Sbjct: 301 RLLDTPLSQSFSHDTIQTITQAVQKDFAITYDQVQQAETKALTSKVFLITGGPGTGKTT 360

15 Query: 361 VINGIIEAYSELHHIDLNKNDIPIVLAAPTGRAARRMNETGLPSATIHRHLGLNGDSY 420
 VI GI++AY+ LH IDL+K D+PI+LAAPTGRAARRMNETGLPSATIHRHLGLNGD+DY
 Sbjct: 361 VIRGILQAYANLHQIDLKDLPIVLAAPTGRAARRMNETGLPSATIHRHLGLNGDNDY 420

20 Query: 421 QSLDDYLDCLSLIIDEFSMVDTWLANQLFDALDSHTQVIIVGSDQLPSVGPQVLADLL 480
 Q+++DYLDCL I+I+DEFSMVDTWLANQL A++S TQVIIVGSDQLPSVGPQVL+DLL
 Sbjct: 421 QAMEDYLDCLLIVDEFSMVDTWLANQLLGAINSTQVIIVGSDQLPSVGPQVLSDLL 480

25 Query: 481 NINALPHVKLEKIFRQSEESTIVTLANQMRQGFLEPDTAKKADRSYFEASANIIPNMIS 540
 +N+LP + L+KIFRQS+ESTIV LA+QMR+G L DF KKADRSYFEA A IP+MI
 Sbjct: 481 KVNSLPQIALQKIFRQSQESTIVNLADQMRGILAADFRDKKADRSYFEAQAAFIPDMIQ 540

30 Query: 541 KIVQSALKSGIEAHEIQILAPMYRGQAGINNLNLMQNLNPLKDNQFTFNDINFRIGD 600
 KIV SA+KSGI A EIQILAPMY+GQAGIN+LN +MQ LNLPL+ +F FND +FR GD
 Sbjct: 541 KIVLSAISKGIPAEIEIQILAPMYKGQAGINHLNQLMQELNPLQGQTEFLFNDTHFRKGD 600

35 Query: 601 KVLHLVNDTELNVFNQDIGYITDLIPAKYTESKQDEIYMTFDGQEVYQRKEWLKITLAY 660
 KVLHLVND +LNVFNQDIGYITDLIPAKYTESKQDE+ + FDG EV Y R EWLK+TLAY
 Sbjct: 601 KVLHLVNDLQNLNVFNQDIGYITDLIPAKYTESKQDELILDFDGSEVTYPRNEWLKITLAY 660

40 Query: 661 AMSIHKSQSGSEFQVVILPITRQSGRMLQRLNIYTAITRSKSKLILLGEIGAFDFAVKNEG 720
 AMSIHKSQSGSEFQVVILPITRQSGR+LQRN+IYTAITRSKSKLILLGE AF++A+K+EG
 Sbjct: 661 AMSIHKSQSGSEFQVVILPITRQSGRLLQRNVIYTAITRSKSKLILLGEYTAFFEYAIKHEG 720

45 Query: 721 AKRNTYLIERFENKQBIANSQKIEDSSIDQKI-----DNTIINTSIPKTATPIEQ 770
 KR TYLIERF+ + ++A+SQ ++ ++ D++ ++S + P E
 Sbjct: 721 DKRQTYLIERFQEQSDLASSQPNQELKSKEQTSLSNTATLEDDSSQSSSQSTNSNPTE 780

Query: 771 TNLISKITYRLTEENYLTIDPMIGINQQDISAIFDSK 806
 + +RLT ENY TID MIG+ + DI+ F K
 Sbjct: 781 SQSDNDDFRITPENYSTIDSMIGLTESDIALFFQKK 816

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1631

45 A DNA sequence (GBSx1726) was identified in *S.agalactiae* <SEQ ID 5037> which encodes the amino acid sequence <SEQ ID 5038>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -8.23 Transmembrane 9 - 25 (7 - 29)
 50 ----- Final Results -----
 bacterial membrane --- Certainty=0.4291(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
 55

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB69116 GB:U90721 signal peptidase I [Streptococcus pneumoniae]
 Identities = 120/201 (59%), Positives = 144/201 (70%), Gaps = 9/201 (4%)
 60 Query: 2 KEFIKEWGVFILILSLFLLSRIFLWQFVKVDGHSMDPTLADKEQLVVLKQTKINRFDIVV 61
 K F+KEWG+F+LILSL LSRIF W V+V+GHSMDPTLAD E L V+K I+RFDIVV
 Sbjct: 5 KNFLKEWGLFLLILSLLLSRIFLWQFVKNVVEGHSMDPTLADGEILFVVKHLPIDRFDIVV 64

-1825-

Query: 62 ANEEEGGQKKKIVKRVIGMPGDVIKYKNDTLTINNKKTEEPYLKEYTKLFKKDKLQEKYS 121
 A+EE+G K IVKRVIGMPGD I+Y+ND L IN+K+T+EPYL +Y K FK DKLQ YS
 Sbjet: 65 AHEEDG--NKDIVKRVIGMPGDTIRYENDKLYINDKETDEPYLADYIKRFKDDKLQSTYS 122

Query: 122 -----YNPLFQDLAQSSSTAFTTDSNGSSEFTTVVPKGHYLVGDDRIVSKDSRAVGPF 174
 F+ +AQ + AFT D N ++ F+ VP+G Y L+GDDR+VS DSR VG F
 Sbjet: 123 GKGFEKNKGTFFRSIAQKAQFTVDVNYNTNFSFTVPEGEYLLLGDDRIVSSDSRHVGTF 182

Query: 175 KKSTIVGEVKFRFWPIRRFGT 195
 K I GE KFRFWPI R GT
 Sbjet: 183 KAKDITGEAKFRFWPITRIGT 203

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5039> which encodes the amino acid sequence <SEQ ID 5040>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.50 Transmembrane 35 - 51 (35 - 51)

----- Final Results -----
 bacterial membrane --- Certainty=0.1999(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9157> which encodes the amino acid sequence <SEQ ID 9158>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty= 0.300(Affirmative) < succ>
 bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 131/197 (66%), Positives = 152/197 (76%)

Query: 1 MKEFIKEWGVFILILSLFLLSRIFLWQFVKVDGHSM DPTLADKEQLVVLKQTKINRFDIV 60
 MK+FIKEWG F L L LF LSR+FLWQ VKVDGHSM DPTLA E+L+V Q +I+RFDIV
 Sbjet: 23 MKQFIKEWGPFTLFLILFGLSRLFLWQAVKVDGHSM DPTLAHGERLIVFNQARIDRFDIV 82

Query: 61 VANE EGGQKKKIVKRVIGMPGDVIKYKNDTLTINNKKTEEPYLKEYTKLFKKDKLQEKY 120
 VA EEE GQKK+IVKRVIG+PGD I Y +DTL IN KKT EPYL EY K FK DKLQ+ Y
 Sbjet: 83 VAQEEENGQKKEIVKRVIGLPGDTISYND DTL YINGKKTVEPYLA EYLKQFKNDKLQKTY 142

Query: 121 SYNPLFQDLAQSSSTAFTTDSNGSSEFTTVVPKGHYLVGDDRIVSKDSRAVGPFKKSTIV 180
 +YN LFQ LA++S AFTT+S G + F VPKG Y L+GDDRIVS+DSR VG FKK ++
 Sbjet: 143 AYN TLFQQLAETSDAFTTNSSEGQTRFEMSVPKGEYLLLGDDRIVSRDSREVGSGFKKENLI 202

Query: 181 GEVKFRFWPIRRFGTIN 197
 GEVK RFWP+ + N
 Sbjet: 203 GEVKARFWPLNKMTVFN 219

SEQ ID 5038 (GBS268) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 4; MW 50.3kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 9; MW 25.3kDa) and in Figure 160 (lane 2-4; MW 25.3kDa).

GBS268-His was purified as shown in Figure 222, lane 8.

-1826-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1632

A DNA sequence (GBSx1727) was identified in *S.agalactiae* <SEQ ID 5041> which encodes the amino acid sequence <SEQ ID 5042>. This protein is predicted to be ribonuclease HIII (rnhB). Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4728(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10119> which encodes amino acid sequence <SEQ ID 10120> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45437 GB:U93576 ribonuclease HII [Streptococcus pneumoniae]
Identities = 176/282 (62%), Positives = 219/282 (77%), Gaps = 13/282 (4%)

Query: 16 EKIRTDLAQHHSNNNNPYVVSFAKISGATVLLYTSGLVFGSGNASHIAQKYGF--IEQK 73
E +T LA + NPY+ + K+ ATV +YTSKG++ QG A A +G+ +EQ
Sbjct: 18 EHYQTS LAP---SKNPYIRYFLKLPQATVSIYTSGLKILQGEAKEYASFFGYQAVEQ- 72

Query: 74 ESCSSESQDIPITIGTDEVGNNGSYFGGLAVVASFVTPKDHAYLKKLGVGDSKTLTDQKIKQ 133
+ Q++P+IGTDEVGNNGSYFGGLAVVA+FVTP H +L+KLGVDGSKTLTDQKI+Q
Sbjct: 73 ----TSGQNLPLIGTDEVGNNGSYFGGLAVVAFAFVTPDQHDFLRKLGVGDSKTLTDQKIRQ 128

Query: 134 IAPLLEKAIPHKALLLSPQKYNQVSPNNKHNAVSVKVALHNQAIFFLLQDGFEPKIV 193
IAP+L++ I H+ALLLSP KYN+V+ +++NAVSVKVALHNQAI+LLLQ G +PEKIVI
Sbjct: 129 IAPILKEKIQHQAALLLSPSKYNEVIG--DRYNAVSVKVALHNQAIYLLQKGVQPEKIVI 186

Query: 194 DAFTSSKNYQNYLKNKKNQFKQTITLEEKAENKYLAVAVSSIIARNLFLENLNKLSDDVG 253
DAFTS+KNY YL E N+F I+LEEKAE KYLAVAVSS+IAR+LFLENL L ++G
Sbjct: 187 DAFTSAKNYDKYLAQETNRFSNPISLEEKAEKYLAVAVSSVIARDLFLENLENLGRELG 246

Query: 254 YKLPSGAGHQSDKVASQLLKAYGISSLEHCAKLHFANTKKAQ 295
Y+LPSGAG SDKVASQ+L+AYG+ L CAKLHF NT+KA+
Sbjct: 247 YQLPSGAGTASDKVASQILQAYGMQGLNFCAKLHFKNTEKAK 288

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5043> which encodes the amino acid sequence <SEQ ID 5044>. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2148(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 194/298 (65%), Positives = 240/298 (80%), Gaps = 2/298 (0%)

Query: 3 MNTIVMQADKKLQEKIRTDLAQHHSNNNNPYVVSFAKISGATVLLYTSGLVFGSGNASH 62
MNT+V++ D L + ++ LA + IS+ N YV F+AK +G TVLLY SGKLV QG+ A+
Sbjct: 1 MNTLVLKIDAILSKHLKKQLAPYTISSQNTYVAFAAKNGVTVLLYKSGKLVLQNGANA 60

-1827-

Query: 63 IAQKYGFIEQKE--SCSSESQDIPIIGTDEVGNGSYFGGLAVVASFVTPKDHAYLKKLGV 120
 +AQ+ K S+ SQDIPIIG+DEVGNGSYFGG+AVVASFV PKDH++LKKLGV
 Sbjct: 61 LAQELNLPVAKTVFEASNNSQDIPIIGSDEVGNGSYFGGIAVVASFVDPKDHSLKKLGV 120

5 Query: 121 GDSKTLTDQIKQIAPLLEKAIPHKALLSPQKYNQVSPNNKHNAVSVKVALHNQAIFL 180
 DSK L+D+ I+QIAPLLEK IPH++LLLSP+KYN++V + +NA+S+KVALHNQAIFL
 Sbjct: 121 DDSKKLSDKTIQQIAPLLEKQIPHQSLLSPKYNELVGKSKPYNALSIKVALHNQAIFL 180

10 Query: 181 LLQDGFEPKIVIDAFTSSKNYQNYLKNEKNQFKQTITLEEKAENKYLAVAVSSIIARNL 240
 LLQ G +P++IVIDAFTS NY+ +LK EKN F +T +EKAE+ YLAVAVSSIIARNL
 Sbjct: 181 LLQKGIQPKQIVIDAFTSQSNYEKHLKKEKNHFPNPLTFQEKAESHYLAVAVSSIIARNL 240

Query: 241 FLENLNKLSDDVGKLPSPGAGHQSDKVASQLLKAYGISSLEHCAKLHFANTKKAQALL 298
 FL+NL++L D+GY+LPSGAG SDKVASQLL AYG+SSLE+ AKLHFANT KAQALL
 15 Sbjct: 241 FLDNLDQLGQDLGYQLPSGAGSASDKVASQLLAAYGMSSLEYSAKLHFANTHKAQALL 298

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1633

20 A DNA sequence (GBSx1728) was identified in *Sagalactiae* <SEQ ID 5045> which encodes the amino acid sequence <SEQ ID 5046>. This protein is predicted to be heat shock protein 70. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3874(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5047> which encodes the amino acid sequence <SEQ ID 5048>. Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3442(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40

An alignment of the GAS and GBS proteins is shown below.

Identities = 65/92 (70%), Positives = 76/92 (81%)

45 Query: 11 NRYKFVFGDKPLTLTLDKDNLFMEIEIERVATEKYEAIKEKLPNADNETTAILMAINLSV 70
 NRYKF FG+K LTLTLDKDNLFMEIE+ERVA EKY+A+K LP AD+ETTAILMAIN LS
 Sbjct: 5 NRYKFTFGEKTLTLTLDKDNLFMEIEVERVAKEKYQALKNHLPEADDETAILMAINTLST 64

50 Query: 71 QLSREIDIEKMEDELNKLRSKTISDIKEKVSE 102
 QLSREI IEKME E+ LR KT+ ++EK ++
 Sbjct: 65 QLSREIAIEKMEAEILDRLQKTLVGLQEKANQ 96

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1828-

Example 1634

A DNA sequence (GBSx1729) was identified in *S.agalactiae* <SEQ ID 5049> which encodes the amino acid sequence <SEQ ID 5050>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 48
      >>> Seems to have an uncleavable N-term signal seq
          INTEGRAL      Likelihood = -10.99      Transmembrane 124 - 140 ( 114 - 148)
          INTEGRAL      Likelihood = -5.84      Transmembrane 22 - 38 ( 21 - 40)
          INTEGRAL      Likelihood = -4.88      Transmembrane 2 - 18 ( 1 - 20)
10      INTEGRAL      Likelihood = -1.97      Transmembrane 84 - 100 ( 84 - 100)

      ----- Final Results -----
          bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

      >GP:BAB06827 GB:AP001517 unknown conserved protein in B. subtilis
      [Bacillus halodurans]
      Identities = 59/182 (32%), Positives = 98/182 (53%), Gaps = 14/182 (7%)
20
      Query: 1  MLSLLLLIIIVHWFYIGYSRGIFLQVFYVLSMSVSLMIASQFYQELASQITLWVPYS--N 58
      MLS++LL I++ F+IG RG+ LQ+ ++L + + +A ++Y +A+ I LW+PY +
      Sbjct: 1  MLSVILLFILLCSFFIGKRRGLILQLVHLLGFVAFFVAYKYYAPVATYIRLWIPYPQFS 60

      Query: 59  PVQGVVEVYFFKDISKFQLSHVYAGVAFVFIY----SLSYLVGRLLGVLLHLAPVEHFDS 114
      P V + I F +VYY+G+AF ++ L ++VG +L L HL +
      Sbjct: 61  PDSPVTML-----IEAFNFENVYYSGLAFALLFIGTKILLHIVGSMDFLTHLPILRSV-- 114

      Query: 115 LQNNIISGFLAVLVCLLFMSMCLTILATVPMSFVQEKLWNSLFVRFLINDLPFFSQFLVR 174
      N + G L + L M + L + A +P+ VQ L SL +F++N PF S+F+
30      Sbjct: 115 --NGWLGGILGFVEVYLIMFVLLYVGALLPIETVQTHLNQSLVAQFIMNHTPFLSEFIRN 172

      Query: 175 TW 176
      W
35      Sbjct: 173 LW 174

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5051> which encodes the amino acid sequence <SEQ ID 5052>. Analysis of this protein sequence reveals the following:

```

40      Possible site: 59
      >>> Seems to have a cleavable N-term signal seq.
          INTEGRAL      Likelihood = -8.17      Transmembrane 124 - 140 ( 117 - 148)
          INTEGRAL      Likelihood = -4.73      Transmembrane 84 - 100 ( 78 - 105)
          INTEGRAL      Likelihood = -0.00      Transmembrane 156 - 172 ( 156 - 172)

45      ----- Final Results -----
          bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

50 The protein has homology with the following sequences in the databases:

```

      >GP:BAB06827 GB:AP001517 unknown conserved protein in B. subtilis
      [Bacillus halodurans]
      Identities = 57/177 (32%), Positives = 98/177 (55%), Gaps = 2/177 (1%)
55
      Query: 1  MLSLLIVLILTWNFYIGYSRGIIQSFYVLGALLSLLVANRFYIGLAHKLTLWIPYSNPV 60
      MLS++++ IL +F+IG RG+ILQ ++LG + + VA ++Y +A + LWIPY
      Sbjct: 1  MLSVILLFILLCSFFIGKRRGLILQLVHLLGFVAFFVAYKYYAPVATYIRLWIPYPQFS 60

      Query: 61  EGTSVFFFKSVDIFVLDKVYVYAGLAFPIIFLLGYALSRLGIFVHFLLLNYFDNQWTKCL 120
      + V ++ F + VYY+G+AF ++F+ L +G + FL L
60      Sbjct: 61  PDSPVTML--IEAFNFENVYYSGLAFALLFIGTKILLHIVGSMDFLTHLPILRSVNGWL 118

```

-1829-

Query: 121 SGGLAFLVSLFLNMLLSIFATVPMPLQHYLHSSFLARLVIEHLPPLTIIQKLWI 177
 G L F+ L + +LL + A +P+ +Q +L+ S +A+ ++ H P L+ I+ LWI
 Sbjct: 119 GGILGFVEVYLIMFVLLYVGALLPIETVQTHLNQSLVAQFIMNHTPFLSEFIRNLWI 175

An alignment of the GAS and GBS proteins is shown below.

Identities = 87/176 (49%), Positives = 123/176 (69%)

Query: 1 MSLLLLLIIIVIHFIYIGYSRGIFLQVVFVLMMSVSLMIASQFYQELASQITLWVPYSNPV 60
 MSLLL+++I+ W+FYIGYSRGI LQ FYVL +++SL++A++FY LA ++TLW+PYSNPV
 Sbjct: 1 MSLLLIVLILTWNFIYIGYSGIILQSFYVLGALLSLLVANRFYIGLAHKLTLWIPYSNPV 60

Query: 61 QGVEVYFFKDISKQFQLSHVYAGVAFVFIYSLSYLVGRLLGVLLHLAPVEHFDSLQNNII 120
 +G V+FFK + F L VYAG+AF I+ L Y + R LG+ +H + +FD+ +
 Sbjct: 61 EGTSVFFFKSVDIFVLDKVYAGLAFFIIFLLGYALSRLGIFVHFLLLNYFDNQWTKCL 120

Query: 121 SGFLAVLVCLLFMSMCLTILATVPMFSVQEKLNWSLFRFLINDLPFFSQFLVRTW 176
 SG LA LV LLF++M L+I ATVPM F+Q L +S R +I LP + + + W
 Sbjct: 121 SGGLAFLVSLFLNMLLSIFATVPMPLQHYLHSSFLARLVIEHLPPLTIIQKLW 176

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1635

A DNA sequence (GBSx1730) was identified in *S.agalactiae* <SEQ ID 5053> which encodes the amino acid sequence <SEQ ID 5054>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4176(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10117> which encodes amino acid sequence <SEQ ID 10118> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14818 GB:Z99118 similar to DNA mismatch repair protein
 [Bacillus subtilis]
 Identities = 320/790 (40%), Positives = 466/790 (58%), Gaps = 18/790 (2%)

Query: 10 MNNKILEQLEFNKVKELILPYLKTEQSQEELSELEPMTEAPKIEKSFNEISDMEQIFVEH 69
 M K+L LEF+KVKE ++ + + +E L EL+P +I+K +E+ + I
 Sbjct: 1 MQQKVLSALEFHKVKEQVIGHAASSLGKEMLLLELKPSASIDEIKQLDEVDEASDIIRLR 60

Query: 70 HSFGIVSLSSISSESLKRLELSADLNIEQLLAIKVLQSSSDMIHFYSDL--DNVSFQSLD 127
 L I +L+R E+ + L+ E I +L+ M HF + + D V +
 Sbjct: 61 GQAPFGGLVDIRGALRRAEIGSVLSPSEFTEISGLLYAVKQMKHFITQMAEDGVDIPLIH 120

Query: 128 RLFENLEQFPNLQGSFQA-INDGGFLEHFASPELERIRRLTNSERRVRQILQDMLKEKA 186
 + E L +L+ + I+D G + AS L IR QL E RVR L+ ML+ +
 Sbjct: 121 QHAEQLITLSDLERDINSCIDDHGEVLDHASETLRGIRTQLRTLERSVRDRLESMLRSSS 180

Query: 187 --ELLSENLIASRSGRSVLVKNYRNIRISGVVHDISSSGSTVYIEPRAVVTLNNEITQL 244
 ++LS+ ++ R+ R V+PVK YR+ G+VHD SSG+T++IEP+A+V +N + Q
 Sbjct: 181 ASKMLSDTIVTIRNDRFVIVPVKQYRSSYGGIVHDTSSSGATLFIEPQAIVDMNNSLQQA 240

Query: 245 RADERHEESRILHAFSDLLRPHVATIRNNAWILGHLDVRAKYLFMNDKATPEISNDS 304
 + E+ E RIL ++ + + + +L LDF+ AK + KAT P +++
 Sbjct: 241 KVKEKQEIERRILRVLTEKTAEYTEELFLDLQVLQTLDFIFAKARYAKAVKATKPIMNDTG 300

-1830-

Query: 305 TLALINVRHPLL--SNPVANDLHFDQDLTAIVITGPNTGGKTI MLKTLGLAQLMGQSGLP 362
 + L RHPLL VAND+ +D + IVITGPNTGGKT+ LKTLGL LM QSGL
 Sbjct: 301 FIRLKKARHPLLPPDQVVANDIELGRDFSTIVITGPNTGGKTVTLKTLGLLTLMAQSGLH 360

Query: 363 VLADKGSKI AVFNIFADIGDEQSIEQSLSTFSSHMTHIVSILNEADHNSLVLFDELGAG 422
 + AD+GS+ AVF ++FADIGDEQSIEQSLSTFSSHM +IV IL + + NSLVLFDELGAG
 Sbjct: 361 IPADGSEAAVFEHV FADIGDEQSIEQSLSTFSSHMVNIVGILEQVNENSLVLFDELGAG 420

Query: 423 TDPQEGASLAMAILEHLRLSNIKTMATTHYPELKAYGIETNFVENASMEFDAETLSPTYR 482
 TDPQEGA+LAM+IL+ + +N + +ATTHYPELKAYG V NAS+EFD ETLSPTY+
 Sbjct: 421 TDPQEGAALAMSILDDVHRTNARVLATTHYPELKAYGYNREGVMNASVEFDIETLSPTYK 480

Query: 483 FMQGVPGRSNAFEIASRLGLAPFIVKQAK-QMTSDSDVNRIIEQLEAQTLETRRRLDHI 541
 + GVPGRSNAFEI+ RLGL I+ QAK +MT ++V+ +I LE L
 Sbjct: 481 LLIGVPGRSNAFEISKRLGLPDHII GQAKSEMTAEHNEVDTMIASLEQSKKRAEEELSET 540

Query: 542 KEVEQENLKFENRAVKKLYNEFSHERDKELEKIYQEAQEIVDMALNESDTILKKL---ND 597
 + + +E K ++ +++ E + ++DK LE+ Q+A E V A+ E++ I+ +L +
 Sbjct: 541 ESIRKEAEKLHKEQQQIIELNSKKDKMLEEAEQQAEEKVKAAMKEAEDIHELRTIKEE 600

Query: 598 KSQLKPHEIIDAKAQIKKLAPQVDLSKNKVLNKAKKIKAAAPRIGDDIIVTSYGQRTL 657
 K HE+I+AK +++ P + SK K +K R + GD++ V ++GQ+GTL
 Sbjct: 601 HKSFKDHELINAKKRLLEGAMPAFEKSKKPEKPKTKQ---RDFKPGDEVKVLTFGQKGTL 656

Query: 658 TSQDKDGRWEAQVGIKMTLTQDEFTLVVRVQEEQKVKSKQINVVKADSSGPRARLDLRG 717
 + W Q+GI+KM + + + ++ E K K K I VK D LDLRG
 Sbjct: 657 LEKTGGNEWNVQIGILKMVKKEKLEFIKSAPEPK-KEKMITAVKGKDYH-VSLELDLRG 714

Query: 718 KRYEEMQELDNFIDQALLNMGQVDIIHGIGTGVIREGVTKYLRRNKHVKHFAYAPQNA 777
 +RYE A+ ++ ++D A+L +V IING GTG +R+GV L+ ++ VK +
 Sbjct: 715 ERYENALSRVEKYLD DAVLAGYPRVSI IINGKGTGALRKGVQDLLKNHRSVKSSRFGEAGE 774

Query: 778 GGSGATIVTL 787
 GGSG T+V L
 Sbjct: 775 GGSGVTIVVEL 784

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5055> which encodes the amino acid sequence <SEQ ID 5056>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3843(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 775/787 (98%), Positives = 781/787 (98%)

Query: 2 INLGIMKSMNNKILEQLEFNKVKELILPYLKTEQSQEELSELEPMTEAPKIEKSFNEISD 61
 I LGIMKSMNNKILEQLEFNKVKEL+LPYLKTEQSQEEL ELEPMTEAPKIEKSFNEISD
 Sbjct: 32 IILGIMKSMNNKILEQLEFNKVKELLLPYLKTEQSQEELLELEPMTEAPKIEKSFNEISD 91

Query: 62 MEQIFVEHHSFGIVSLSSISSESLKRLELSADLNIEQLLAIKKVLQSSSDMIHFYSDLDNV 121
 MEQIFVEHHSFGIVSLSSISSESLKRLELS DLNIEQLLAIKKVLQSSSDMIHFYSDLDNV
 Sbjct: 92 MEQIFVEHHSFGIVSLSSISSESLKRLELSTDLNIEQLLAIKKVLQSSSDMIHFYSDLDNV 151

Query: 122 SFQSLDRLFENLEQFPNLQGSFQAINDGGFLEHFASPELERIRRQLTNSERRVRQILQDM 181
 SFQSLDRLFENLEQFPNLQGSFQAINDGGFLEHFASPELERIRRQLTNSERRVRQILQDM
 Sbjct: 152 SFQSLDRLFENLEQFPNLQGSFQAINDGGFLEHFASPELERIRRQLTNSERRVRQILQDM 211

Query: 182 LKEKAELLENLIASRSGRSVLPVKNTYRNRISGVVHDISSSGSTVYIEPRAVVTLINEEI 241
 LKEKAELLENLIASRSGRSVLPVKNTYRNRISGVVHDISSSGSTVYIEPRAVVTLINEEI
 Sbjct: 212 LKEKAELLENLIASRSGRSVLPVKNTYRNRISGVVHDISSSGSTVYIEPRAVVTLINEEI 271

-1831-

5
Query: 242 TQLRADERHEESRILHAFSDLLRPHVATIRNNAWILGHLDVFRAKYLFMSDNKATIP+IS 301
TQLRADERHEE RILHAFSDLLRPHVATIRNNAWILGHLDVFRAKYLFMSDNKATIP+IS
Sbjct: 272 TQLRADERHEEGRILHAFSDLLRPHVATIRNNAWILGHLDVFRAKYLFMSDNKATIPKIS 331

10
Query: 302 NDSTLALINVRHPLLSNPVANDLHFDQDLTAIVITGPNTGGKTI MLKTLGLAQLMGQSG 361
NDSTLALINVRHPLLSNPVANDLHFD DLTAVITGPNTGGKTI MLKTLGLAQLMGQSG
Sbjct: 332 NDSTLALINVRHPLLSNPVANDLHFDHDLTAIVITGPNTGGKTI MLKTLGLAQLMGQSG 391

15
Query: 422 GTDPQEGASLAMAILEHLRLSNIKTMTTHYPELKAYGIETNFVENASMEFDAETLSPTY 481
GTDPOEGASLAMAILEHLRLS+IKTMTTHYPELKAYGIETNFVENASMEFDAETLSPTY
Sbjct: 452 GTDPQEGASLAMAILEHLRLSHIKTMTTHYPELKAYGIETNFVENASMEFDAETLSPTY 511

20
Query: 482 RFMQGVPGRSNAFEIASRLGLAPFIVKQAKQMTSDSDVNRIIEQLEAQTLETRRLDHI 541
RFMQGVPGRSNAFEIASRLGLAPFIVKQAKQMTSDSDVNRIIEQLEAQTLETRRLDHI
Sbjct: 512 RFMQGVPGRSNAFEIASRLGLAPFIVKQAKQMTSDSDVNRIIEQLEAQTLETRRLDHI 571

25
Query: 542 KEVEQENLKFNRAVKKLYNEFSHERDKELEKIYQEAQEIVDMALNESDTILKKLNDKSQL 601
KEVEQENLKFNRAVKKLYNEFSHERDKELEKIYQEAQEIVDMALNESDTILKKLNDKSQL
Sbjct: 572 KEVEQENLKFNRAVKKLYNEFSHERDKELEKIYQEAQEIVDMALNESDTILKKLNDKSQL 631

30
Query: 602 KPHEIIDAKAQIKKLAPQVDLSKNKVLNKAKKIKAARAPRIGDDIIVTSYQGRGTLTSQL 661
KPHEIIDAKAQIKKLAPQVDLSKNKVLNKAKKIKAARAPRIGDDIIVTSYQGRGTLTSQL
Sbjct: 632 KPHEIIDAKAQIKKLAPQVDLSKNKVLNKAKKIKAARAPRIGDDIIVTSYQGRGTLTSQL 691

35
Query: 662 KDGRWEAQVGIKMTLTQDEFTLVRVQEEQVKSKQINNVKKADSSGPRARLDLRGKRYE 721
KDGRWEAQVGIKMTLTQDEF+LVRVQEEQVK+KQINNVKKAD SGPRARLDLRGKRYE
Sbjct: 692 KDGRWEAQVGIKMTLTQDEFSLVRVQEEQVKNKQINNVKKADSGGPRARLDLRGKRYE 751

40
Query: 722 EAMQELDNFIDQALLNNMGQVDIIHGIGTGVIREGVTKYLRNKHVKHFAYAPQNAGGSG 781
EAMQELD+FDQALLNNMGQVDIIHGIGTGVIREGVTKYLRNKHVKHFAYAPQNAGGSG
Sbjct: 752 EAMQELDHFDQALLNNMGQVDIIHGIGTGVIREGVTKYLRNKHVKHFAYAPQNAGGSG 811

Query: 782 ATIVITLG 788
ATIVITLG
Sbjct: 812 ATIVITLG 818

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1636

45 A DNA sequence (GBSx1731) was identified in *S.agalactiae* <SEQ ID 5057> which encodes the amino acid sequence <SEQ ID 5058>. This protein is predicted to be thioredoxin (trxA). Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2721(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 10115> which encodes amino acid sequence <SEQ ID 10116> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB40815 GB:AJ133006 thioredoxin [*Listeria monocytogenes*] (ver

-1832-

2)

Identities = 64/100 (64%), Positives = 78/100 (78%), Gaps = 1/100 (1%)

Query: 15 MALEVTDATFVEETKEGLVLIDFWATWCGPCRMQAPILEQLSQEIDEDELKILKMDVDEN 74
 M E+TDATF +ET EGLVL DFWATWCGPCRM AP+LE++ +E E LKI+KMDVDEN
 Sbjct: 1 MVKEITDATFQEETSEGLVLIDFWATWCGPCRMVAPVLEEIQEERGE-ALKIVKMDVDEN 59

Query: 75 PETARQFGIMSIPTLMFKKDGEVVKQVAGVHTKDQLKAI 114
 PET FG+MSIPTL+ KKDGEVV+ + G K++L +I
 Sbjct: 60 PETPGSFGVMSIPTLLIKKDGEVVETIIGYRPKEELDEVI 99

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5059> which encodes the amino acid sequence <SEQ ID 5060>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2721(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1637

A DNA sequence (GBSx1732) was identified in *S.galactiae* <SEQ ID 5061> which encodes the amino acid sequence <SEQ ID 5062>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -7.54 Transmembrane 170 - 186 (167 - 191)
 INTEGRAL Likelihood = -5.52 Transmembrane 87 - 103 (86 - 107)
 INTEGRAL Likelihood = -4.62 Transmembrane 105 - 121 (104 - 126)

----- Final Results -----

bacterial membrane --- Certainty=0.4015(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA60798 GB:X87369 ORF3 [Clostridium perfringens]

Identities = 27/67 (40%), Positives = 52/67 (77%)

Query: 1 MEIGQQIIRYRKQQAALSQEELAEKVYVSQRSISNWENDKTYPDHSLLLLSQIFQVSLDQ 60
 M++ +++ RK++ LSQE+LAEK+ +SRQ++S WE+ ++ PD++ L++LS+++ V++D
 Sbjct: 1 MKLAEKLQLMRKREGLSQEDLAELGISRQAVSKWESGQVDPDLNKLIIILSELYNVTIDY 60

Query: 61 LIKGDIE 67

L+K E

Sbjct: 61 LVKETYE 67

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1739> which encodes the amino acid sequence <SEQ ID 1740>. Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -8.86 Transmembrane 173 - 189 (169 - 194)
 INTEGRAL Likelihood = -5.52 Transmembrane 90 - 106 (89 - 110)
 INTEGRAL Likelihood = -4.62 Transmembrane 108 - 124 (107 - 129)

----- Final Results -----

-1833-

bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 187/195 (95%), Positives = 191/195 (97%)

Query: 1 MEIGQQIIRYRKQALSQEELAEKVYVSRQISNWENDKTPDIHSLLLLSQIFQVSLDQ 60
 MEIGQQIIRYRKQALSQE+LAEKVYVSRQISNWENDKTPDIHSLLLLSQIFQVSLDQ
 10 Sbjet: 4 MEIGQQIIRYRKQALSQEELAEKVYVSRQISNWENDKTPDIHSLLLLSQIFQVSLDQ 63

Query: 61 LIKGDIEKMKYTITQVDKKNFERTKVMVTLMILLMISSYPLVYFLEWLGLGIFVLSII 120
 LIKGDIEKMKYTITQVDKKNF+RDTKVMVTLMILLMISSYPLVYFLEWLGLGIFVLSII
 15 Sbjet: 64 LIKGDIEKMKYTITQVDKKNFRTKVMVTLMILLMISSYPLVYFLEWLGLGIFVLSII 123

Query: 121 TMTYANRVERFKKKYDVQTYKEILAVSSGKLLDEIEKREERAKLPYQKPLIVTVFFLITV 180
 TMTYANRVERFKKKYDVQ YKEILAVS+GKLLDEIEKREERA LPYQKPLIVTVFFLITV
 20 Sbjet: 124 TMTYANRVERFKKKYDVQPYKEILAVSNGKLLDEIEKREERATLPYQKPLIVTVFFLITV 183

Query: 181 ATFFASRFIFTWLFH 195
 A FASRF+FTWLFH
 25 Sbjet: 184 AFASFASRFMTWLFH 198

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 1638

A DNA sequence (GBSx1733) was identified in *S.agalactiae* <SEQ ID 5063> which encodes the amino acid sequence <SEQ ID 5064>. This protein is predicted to be adenine glycosylase (mutY). Analysis of this protein sequence reveals the following:

30 Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2385(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9425> which encodes amino acid sequence <SEQ ID 9426> was also identified.

40 The protein has homology with the following sequences in the GENPEPT database.

>GP: BAB04650 GB: AP001510 adenine glycosylase [Bacillus halodurans]
 Identities = 130/331 (39%), Positives = 190/331 (57%), Gaps = 15/331 (4%)

45 Query: 1 MLQQTQVNTVIPYKRFLEWFPQIKDLADAPEEQLLKAWEGGLGYYSRVRNMQKAAQQVMV 60
 MLQQT+V+TVIPYY+ F+ FP ++ LA A E+Q+LKAWEGGLGYYSR RN+Q A ++V+
 Sbjet: 45 MLQQTRVDTVIPYYQAFMRQFPPTLETLAYAEEDQVLKAWEGGLGYYSRARNLQSAVREVVE 104

Query: 61 DFGGIFPHTYDDIASLKGIGPYTAGAIASISFNLPEPAVDGNVMRVMARLFEVNYDIGDP 120
 +GG P T +I+ LKG+GPYTAGAI SI+++ PEPAVDGNVMRV++R+ + DI
 50 Sbjet: 105 SYGGEVPSTRKEISKLGKGVGPYTAGAILSLAYDQPEPAVDGNVMRVLSRVLYIEEDIAKV 164

Query: 121 KNRKIFQAIMEILIDPDRPGDFNQALMDLGTDIESAKTTPRDESPIRFFNAAYLNGTYSK 180
 K R +F++++ LI + P FNQ LM+LG + + +P P+R A+ G +
 55 Sbjet: 165 KTRTLFESLLYDLISKENPSFFNQGLMELGALVCTPTSPGCLLCPVRDHCRFAAGVQEQ 224

Query: 181 YPIKNTKKKPKPMRIQAFVIRNQNQYLLKNTKGRLGGFWSFPPIIETSPILSQQLDLFD 240
 PIK KKKPK ++ A VIRN+ GQ L+E+ + LL W FP +E L
 Sbjet: 225 LPIKAKKKKPKAKQLIAAVIRNEKGQVLIERRPEKGLLAKLWQFPNVE-----LES 275

-1834-

Query: 241 DNQSNPIIWQTQNETFQREYQLKPQWTDNHFPNIKHTFSSHQKWTIELIEGVVKAT-DLPN 299
 + ++ +E F + + + ++H FSH W I + E VK L +
 Sbjct: 276 TKNAQQVLGDYIHERFHLDAAV-----GEYVQTVEHVFSHLIWNIRVYEATVKGVP SLND 330

Query: 300 APHLKWVAIEDFSLYPFATPQKKMLETYLKQ 330
 WV Y F +K+++ L++
 Sbjct: 331 KYEADWVDDRTIENYAFVPVSHQKIIQGNLRK 361

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5065> which encodes the amino acid sequence <SEQ ID 5066>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3579(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 330/333 (99%), Positives = 331/333 (99%)

- Query: 1 MLQQTQVNTVIPYKRFLEWFPQIKDLADAPEEQLLKAWEGLYYSRVRNMQKAAQQVMV 60
 MLQQTQVNTVIPYKRFLEWFPQIKDLADAPEEQLLKAWEGLYYSRVRNMQKAAQQVMV
 25 Sbjct: 52 MLQQTQVNTVIPYKRFLEWFPQIKDLADAPEEQLLKAWEGLYYSRVRNMQKAAQQVMV 111
- Query: 61 DFGGIFPHTYDDIASLKGIGPYTAGAIASISFNLPEPAVDGNVMRVMARLFEVNYDIGDP 120
 DFGGIFPHTYDDIASLKGIGPYTAGAIASISFNLPEPAVDGNVMRVMARLFEVNYDIGDP
 30 Sbjct: 112 DFGGIFPHTYDDIASLKGIGPYTAGAIASISFNLPEPAVDGNVMRVMARLFEVNYDIGDP 171
- Query: 121 KNRKIFQAIMEILIDPDRPGDFNQALMDLGTDIESAKTPRPDESPIRFFNAAYLNGTYSK 180
 KNRKIFQAIMEILIDPDRPGDFNQALMDLGTDIESAKTPRPDESPIRFFNAAYLNGTY K
 35 Sbjct: 172 KNRKIFQAIMEILIDPDRPGDFNQALMDLGTDIESAKTPRPDESPIRFFNAAYLNGTYGK 231
- Query: 181 YPIKNKKPKPMRIQAFVIRNQNQGQYLLEKNTKGRLGGSFPIIETSPLSQQLDLFD 240
 YPIKN KKKPKPMRIQAFVIRNQNQGQYLLEKNTKGRLGGSFPIIETSPLSQQLDLFD
 40 Sbjct: 232 YPIKNPKKKPKPMRIQAFVIRNQNQGQYLLEKNTKGRLGGSFPIIETSPLSQQLDLFD 291
- Query: 241 DNQSNPIIWQTQNETFQREYQLKPQWTDNHFPNIKHTFSSHQKWTIELIEGVVKATDLPNA 300
 DNQSNPIIWQTQNETF+REYQLKPQWTDNHFPNIKHTFSSHQKWTIELIEGVVKATDLPNA
 45 Sbjct: 292 DNQSNPIIWQTQNETFEREYQLKPQWTDNHFPNIKHTFSSHQKWTIELIEGVVKATDLPNA 351
- Query: 301 PHLKWVAIEDFSLYPFATPQKKMLETYLKQKNA 333
 PHLKWVAIEDFSLYPFATPQKKMLETYLKQKNA
 45 Sbjct: 352 PHLKWVAIEDFSLYPFATPQKKMLETYLKQKNA 384

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1639

- 50 A DNA sequence (GBSx1734) was identified in *S.agalactiae* <SEQ ID 5067> which encodes the amino acid sequence <SEQ ID 5068>. This protein is predicted to be maltose/maltodextrin transport system (malG). Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have an uncleavable N-term signal seq

55 INTEGRAL Likelihood = -10.30 Transmembrane 14 - 30 (5 - 35)
 INTEGRAL Likelihood = -6.95 Transmembrane 248 - 264 (242 - 267)
 INTEGRAL Likelihood = -5.15 Transmembrane 75 - 91 (74 - 94)
 INTEGRAL Likelihood = -3.19 Transmembrane 110 - 126 (110 - 127)
 INTEGRAL Likelihood = -2.13 Transmembrane 141 - 157 (138 - 157)

-1835-

INTEGRAL Likelihood = -0.32 Transmembrane 188 - 204 (188 - 204)

----- Final Results -----

5 bacterial membrane --- Certainty=0.5118(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:BA06643 GB:AP001517 maltose/maltodextrin transport system
(permease) [Bacillus halodurans]
Identities = 117/281 (41%), Positives = 169/281 (59%), Gaps = 5/281 (1%)

Query: 1 MNKK--KRLNLTFFVYILLIVLSIMWLFPIVWVLTFRGEGSAFVNYFIPKWTWLDNYAK 58
15 Sbjct: 1 MNKK RL +T +Y+ L+V+ I+ L+P++W V S S F + IP+T + +Y
MNKKVKSRLEVTAIYLFLLVMGIVILYPLLWTVGLSLNPGTSLFSSRMIPETISFRHYEW 60

Query: 59 LFTQNTFFPGQWFLNLTFLVATCTCILSTLITVAMAYSLSRIFKHRNGFLKLALVLNMFP 118
20 Sbjct: 61 LFFDPRSNYLQWYKNTLIVASVTSVCSTFLVALTAYAFSRYRFVGRTYGLYGFLLQLMFP 120

Query: 119 GFMSMIAVYIILKALNLDQTLTALIFVY-SAGAALTFFYIAKGFFDTIPYSLDESAMIDGA 177
25 Sbjct: 121 VLMAMVALYILLNTVNLDTLLGLILYVGTSLIPMNAFLVKGYFDTIPREDESADKLDGA 180

Query: 178 TRLDIFLKITLPLSKPIIVYITALIAFMGPWMDFIKAVILGDATSKYTVAGLFSMLQD 237
25 Sbjct: 181 GHFRIFFTIMLPLAKPILAVVAFNFMSPFMDFILPRIIL-RSPENYTLALGLFNFNVDQ 239

Query: 238 TINQWMSFTAGSVIIAIPITILEFMFMQYYVEGITGGSVK 278
30 Sbjct: 240 FANN-FTRFAAGAILIAIPIATVFLFLQRYLISGLTTGATK 279

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5069> which encodes the amino acid sequence <SEQ ID 5070>. Analysis of this protein sequence reveals the following:

35 Possible site: 39
>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -6.42 Transmembrane 76 - 92 (71 - 97)
INTEGRAL Likelihood = -6.05 Transmembrane 248 - 264 (242 - 267)
INTEGRAL Likelihood = -3.50 Transmembrane 110 - 126 (110 - 127)
40 INTEGRAL Likelihood = -1.33 Transmembrane 129 - 145 (129 - 145)
INTEGRAL Likelihood = -1.33 Transmembrane 188 - 204 (188 - 204)

----- Final Results -----

45 bacterial membrane --- Certainty=0.3569(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

50 >GP:CAA60006 GB:X86014 cymG [Klebsiella oxytoca]
Identities = 119/270 (44%), Positives = 172/270 (63%), Gaps = 7/270 (2%)

Query: 11 LVYATLILSLIIWLFPIAWVILTFRSEGTAYVNYFIPKFTTLNHYINLFTNETFFPGKW 70
55 Sbjct: 12 LVYLFLLLNALVVLGPIVITVMSSSLKPGNNLFSSGFTSFTLEHYHNLLTGT--PYLKW 69

Query: 71 FMNLTIVATFTCIISTFITVAIAYSLSRIKFKFRNGFLKLALILNMFPGFMSMIAIYYIL 130
Sbjct: 70 YKNTFILATCNMLISLVVVTITAFIFSRYRFKAKKKILMSILVLQMFPAFLSMTAIYILL 129

60 Query: 131 KALGLTQTLTALVLVYSSGAALGF--YIAKGFFDTIPYSLDESAMIDGATRMIDFFKITL 188
+ L T L+LVY +G+ L F ++ KG+FD IP SLDE+A IDGA + IFF+I L
Sbjct: 130 SKMNLIDTYIGLLLVYVTGS-LPFMTWLKGYFDAIPTSLDEAKIDGAGHLTIFFEIL 188

Query: 189 PLAKPIIVYTALLAFMGPIWIDFIFAQVILGDATSKYTVAGLFSMLQPDITINNWFMAFTA 248

-1836-

PLAKPI+V+ AL++F GPW+DFI +IL + K T+AIG+FS + ++ N F F A
 Sbjet: 189 PLAKPILVFVALVSFTGPWMDFILPTLIL-RSEDKMTLAIGIFSWISSNSAEN-FTLFAA 246
 Query: 249 GSVLIAVPITLLFMFMQKYYVEGITGGSVK 278
 G++L+AVPITLLF+ QK+ G+ G+VK
 Sbjet: 247 GALLVAVPITLLFIVTQKHITTLGVSFAVK 276

An alignment of the GAS and GBS proteins is shown below.

Identities = 227/278 (81%), Positives = 253/278 (90%)

Query: 1 MNKKRLNLTFFVYIILLIVLSIMMLFPPIVWVLTSTFRGEGSAFVNYFIPKTIWTLNDAKLF 60
 M K+R L VY LI+LSI+WLFP I WV+LTSFR EG+A+VNYFIPKT+TL++Y LF
 Sbjet: 1 MNKRRFQGLVYATLILSIWLFP IAWVILTSTFRSEGTAYVNYFIPKTFTLNHYINLF 60
 Query: 61 TQNTFFPGQWFLNTLFVATCTCILSTLITVAMAYSLSRIFKFRNGFLKLALVLMFPFGF 120
 T TFFPG+WF+NTL VAT TCI+ST ITVA+AYSLSRIFK RNFGLKLAL+LNMFPFGF
 Sbjet: 61 TNETFFPGKWFMTLIVATFTCIISTFITVAIAYSLSRIFKFRNGFLKLALILNMFPFGF 120
 Query: 121 MSMIAVYIILKALNLDQTLTALIFVYSAGAALTFFYIAKGFDDTIPYSLDESAMIDGATRL 180
 MSMIA+YIILKAL L QTLTAL+ VYS+GAAL FYIAKGFDDTIPYSLDESAMIDGATR+
 Sbjet: 121 MSMIAIYIILKALGLTQTLTALVLVYSSGAALGFYIAKGFDDTIPYSLDESAMIDGATRM 180
 Query: 181 DIFLKITLPLSKPIIYVYALIAFMGPWMDFIFAKVILGDATSKYTVAIGLFSMLQDDTIN 240
 DIF KITLPL+KPIIYVYAL+AFMGPW+DFIFA+VILGDATSKYTVAIGLFSMLQ DTIN
 Sbjet: 181 DIFFKITLPLAKPIIYVYALIAFMGPWIDFIFAQVILGDATSKYTVAIGLFSMLQDDTIN 240
 Query: 241 QWFMSFTAGSVIIAIPITILFMFMQKYYVEGITGGSVK 278
 WFM+FTAGSV+IA+PIT+LFMFMQKYYVEGITGGSVK
 Sbjet: 241 NWFMAFTAGSVLIAVPITLLFMFMQKYYVEGITGGSVK 278

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1640

A DNA sequence (GBSx1735) was identified in *S.agalactiae* <SEQ ID 5071> which encodes the amino acid sequence <SEQ ID 5072>. This protein is predicted to be cymF protein (malF). Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.46	Transmembrane	427 - 443 (417 - 447)
INTEGRAL	Likelihood = -10.24	Transmembrane	99 - 115 (96 - 121)
INTEGRAL	Likelihood = -9.39	Transmembrane	166 - 182 (154 - 185)
INTEGRAL	Likelihood = -6.21	Transmembrane	259 - 275 (257 - 276)
INTEGRAL	Likelihood = -6.21	Transmembrane	229 - 245 (223 - 247)
INTEGRAL	Likelihood = -6.10	Transmembrane	44 - 60 (40 - 66)
INTEGRAL	Likelihood = -4.51	Transmembrane	314 - 330 (312 - 331)

----- Final Results -----

bacterial membrane	---	Certainty=0.5585(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA60005 GB:X86014 cymF [Klebsiella oxytoca]

Identities = 174/428 (40%), Positives = 263/428 (60%), Gaps = 21/428 (4%)

Query: 27 SFLIMGLANLKNKQIVKGLLFLISEILFLITFVYQVIPAVKGLISLGTQEQGMFTKTVDG 86
 SFLIMG L + +KG +FL+ +I+ +I+ + ++ A +GLI+LGT Q T G
 Sbjet: 15 SFLIMGATQLISGHWIKGSVFLFLQIV-VISNINLLLNATQGLITLGTVAQ-----TRSG 68
 Query: 87 IKIQVATQGDNSMLMLIFGLASLIFCCVFAYIYWSNIKSAHLLTLKEEGREIPSFKKDI 146

-1837-

I GDNS+ ML+ G+ + IF ++YW NIK A + SF + +
 Sbjct: 69 FDI---VAGDNSIFMLVEGVVAFIFLFFSIFVYWLNIKDAQVCEKCHQ-----SFTEQL 119
 Query: 147 KSLTDGRFHMFLMSIPLIGVLLFTILPLVYMICLAFITNYDH-NHLPPKSLFDWVGANFG 205
 +++ D RF +++ I + F I+P++ + ++ TNY +H+PPK+L DWVG NF
 Sbjct: 120 RTIYDNRFATIMLAPAFIACIAFIIMPMIITVLVSLTNYSAPHHIPKKNLVDWVGLKNFI 179
 Query: 206 NIFSGRMAS-TFFPVLSTLIWAVFATVTNFFFGIILALLINTKGLKFKKMWRITFVITM 264
 +F R+ S TF + WT++WA FAT+ FG +LAL + K + KK WR +F++
 Sbjct: 180 TLFELRIWKTFFVGIGVWTVLWFAFFATLCTCSFGFLALALENKKIIAKKAWRVVFILPY 239
 Query: 265 AVPQFISLLIMRNLLSDAGPVNALLIKWGLISSAHPLPFLSDPVWAKFSIIFVNMWVGIP 324
 A+P F++LLI R LL+ GPNV+ L WG+ S + FLSDP+ AK ++I V++WVG P
 Sbjct: 240 AIPAFVTLIFRLLNGIGPVNSTLNSWGIDS---IGFLSDPLIAKMTVIAVSVWVGAP 295
 Query: 325 VTMLVATGIIMNLPAEQIEAAEIDGANKFQVFQSIITFPQILLIMTPTLIQQFIGNINNFN 384
 ML+ TG + N+P + EA+E+DGA+KFQ F+ IT P +L + P+L+ F N NNF
 Sbjct: 296 YFMLLITGAMTNIPRDLYEASEVDGASKFQQFREITLPMVLHQVAPSLVMTFAHNFNNFG 355
 Query: 385 VIYLLTQGGPTNSTYYQAGSTDLLVTWLYNLTVTAADYNLASVVGILIFILSAVFSLLAY 444
 IYLLT+GGP N Y AG TD+L+TW+Y LT+ Y +ASV+ I+IF+ ++F++ +
 Sbjct: 356 AIYLLTEGGPINPEYRFAGHTDILITWIYKLTLDQYQIASVISIIIFLFLSIFAIWQF 415
 Query: 445 TRINSYKE 452
 R S+KE
 Sbjct: 416 RRMKSFK 423

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5073> which encodes the amino acid sequence <SEQ ID 5074>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -10.93 Transmembrane 98 - 114 (95 - 122)
 INTEGRAL Likelihood = -9.55 Transmembrane 165 - 181 (152 - 184)
 INTEGRAL Likelihood = -9.24 Transmembrane 424 - 440 (419 - 443)
 INTEGRAL Likelihood = -7.91 Transmembrane 43 - 59 (39 - 71)
 INTEGRAL Likelihood = -7.59 Transmembrane 258 - 274 (256 - 275)
 INTEGRAL Likelihood = -6.21 Transmembrane 228 - 244 (222 - 246)
 INTEGRAL Likelihood = -4.09 Transmembrane 311 - 327 (309 - 328)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5373(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA60005 GB:X86014 cymF [Klebsiella oxytoca]
 Identities = 179/426 (42%), Positives = 266/426 (62%), Gaps = 19/426 (4%)
 Query: 26 SSIIMGFANFANKQFIKILFLISELIFLVAFVSQIIPAIRGLVTLGTQTQGMTTKTIDG 85
 S +IMG + +IKG +FL+ +++ ++ ++ ++ A +GL+TLGT Q T G
 Sbjct: 15 SFLIMGATQLISGHWIKGSVFLLFQIV-VISNINLLLNATQGLITLGTVAQ-----TRSG 68
 Query: 86 INIQVAVDGDNSMLMLIFGLASLIFCLVFAYIYWCNLKSARNLYLFKQKGQKIPSPKEDL 145
 +I V GDNS+ ML+ G+ + IF ++YW N+K A+ Q SF E L
 Sbjct: 69 FDI---VAGDNSIFMLVEGVVAFIFLFFSIFVYWLNIKDAQVCEKCHQ-----SFTEQL 119
 Query: 146 ATLTLNGRFHMFLMAIPLIGVLLFTILPLIYMICLAFITNFDH-NHLPPKSLFDWVGLANFG 204
 T+ + RF ++A I + F I+P+I + ++ TN+ +H+PPK+L DWVGL NF
 Sbjct: 120 RTIYDNRFATIMLAPAFIACIAFIIMPMIITVLVSLTNYSAPHHIPKKNLVDWVGLKNFI 179
 Query: 205 NVLSGRM-AGTFFPIFSWTLIWAVFATVTNFFFGIILALLINTKGLKWKKMWRITFVITI 263
 + R+ + TF I WT++WA FAT+ FG +LAL + K + KK WR +F++
 Sbjct: 180 TLFELRIWKTFFVGIGVWTVLWFAFFATLCTCSFGFLALALENKKIIAKKAWRVVFILPY 239
 Query: 264 AVPQFISLLIMRNLLNDEGPNALLNKIGLINGSPLPFLSDPLWAKFSIIFVNMWIGIPFT 323

```

A+P F++LLI R LLN GP+N+ LN G+ S+ FLSDPL AK ++I V++W+G P+
Sbjct: 240 AIPAFVTLILIFRLLLNGIGIPVNSTLNSWGI--DSIGFLSDPLIAKMTVIAVSVWVGAPYF 297

Query: 324 MLIATGIIMNLPSEQIEAAEIDGASKFQVFKSITFPQILLIMTPNLIIQQFIGNINNFNVI 383
ML+ TG + N+P + EA+E+DGASKFQ F+ IT P +L + P+L+ F N NNF I
Sbjct: 298 MLLITGAMTNI PRDLYEASEVDGASKFQQFREITLPMVLHQVAPSLVMTFAHNFNNFGAI 357

Query: 384 YLLTGGGPTNSEYYQAGTDLTLVTWLYKLTVTAAADYNLASVIGILITVSAIFSLLAYTR 443
YLLT GGP N EY AG TD+L+IW+YKLT+ Y +ASVI I+IF +IF++ + R
Sbjct: 358 YLLTEGGPINPEYRFAGHTDILITWIYKLTLDLDFQQYQIASVISIIFI FLFLSIFAIWQFRR 417

Query: 444 TASYKE 449
S+KE
Sbjct: 418 MKSFKE 423

```

Identities = 357/446 (80%), Positives = 404/446 (90%), Gaps = 2/446 (0%)

Query:	11	MSLKEVFKQGDLATKLKSLFLIMGLANLKNKQIVKGLLFLISEILFLITFVYQVIPAVKGLI	70
		+S+ E ++G KLS +IMG AN NKQ +KG+LFLISE++FL+ FV Q+IPA++GL+	
Sbjct:	10	ISVIEALKRGSWDIKLSSIIMGFANFANKQFIKGILFLISELIFLVAFVSQIIPAIRGLV	69
Query:	71	SLGTQEQGMMTTKTVDGIKIQVATQGDNSMLMLIFGLASLIFCCVFAYIYWSNIKSAHLL	130
		+LGTQ QGMMTTKT+DGI IQVA GDNSMLMLIFGLASLIFC VFAYIYW N+KSA +L	
Sbjct:	70	TLGTQTQGMTTKTIDGINIQVAVDGDNSMLMLIFGLASLIFCLVFAYIYWCNLKSARNLY	129
Query:	131	TLKEEGREIPSFKKDIKSLTDGRFHMTLMSIPLIGVLLFTILPLVYMICLAFTNYDHNHL	190
		K++G++IPSFK+D+ +LT+GRFHMTLM+IPLIGVLLFTILPL+YMICLAFTN+DHNHL	
Sbjct:	130	LFKQKGQKIPSFKEDLATLTNGRFHMTLMAIPLIGVLLFTILPLIYMICLAFTNFDHNHL	189
Query:	191	PPKSLFDWVG FANFGNIFSGRMASTFFPVLSWTLI WAVFATVTNFFFGIILALLINTKGL	250
		PPKSLFDWVG ANFGN+ SGRMA TFFP+ SWTLI WAVFATVTNFFFGIILALLINTKGL	
Sbjct:	190	PPKSLFDWVGLANFGNVLSGRMAGTFFPIFSWTLI WAVFATVTNFFFGIILALLINTKGL	249
Query:	251	KFKKMWRITFVITMAVPQFISLLIMRNLLSDAGPVNALLIKWGLISSAHPLPFLSDPVWA	310
		K+KKMWRITFVIT+AVPQFISLLIMRNLL+D GP+NALL K GLI+ + LPFLSDP+WA	
Sbjct:	250	KWKMKWRITFVITIAVPQFISLLIMRNLLNDEGPLNALLNKIGLINGS--LPFLSDPLWA	307
Query:	311	KFSIIFVNMWVGIPVMTLVATGIIMNLP AEQIEAAEIDGANKQVFQSFITFPQILLIMTP	370
		KFSIIFVNMW+GIP TML+ATGIIMNLP+EQIEAAEIDGA+KQVF+SITFPQILLIMTP	
Sbjct:	308	KFSIIFVNMWIGIPFTMLIATGIIMNLPSEQIEAAEIDGASKQVFVKSFITFPQILLIMTP	367
Query:	371	TLIQQFIGNINNFNVIYLLTQG GPTNSTYYQAGSTDLLVTWLYNLTVTAADYNLASVVG	430
		LIQQFIGNINNFNVIYLLT G GPTNS YYQAG+TDLLVTWLY LTVTAADYNLASV+GI	
Sbjct:	368	NLIQQFIGNINNFNVIYLLTG GPTNSEYYQAGT+TDLLVTWLYKLTVTAADYNLASVIG	427
Query:	431	LIFILSAVFSLLAYTRTNSYKEGA AK 456	
		LIF +SA+FSLLAYTRT SYKEGA AK	
Sbjct:	428	LIFTVSAIFSLLAYTRTASYKEGA AK 453	

Lipop: Possible site: -1 Crend: 8
McG: Discrim Score: -12.73
GvH: Signal Score (-7.5): -6.04
 Possible site: 36

```
>>> Seems to have no N-terminal signal sequence
```

ALOM program count: 7 value: -11.46 threshold: 0.0

INTEGRAL	Likelihood	= -11.46	Transmembrane	427 - 443 (417 - 447)
INTEGRAL	Likelihood	= -9.87	Transmembrane	99 - 115 (96 - 121)
INTEGRAL	Likelihood	= -9.39	Transmembrane	166 - 182 (154 - 185)
INTEGRAL	Likelihood	= -6.21	Transmembrane	259 - 275 (257 - 276)
INTEGRAL	Likelihood	= -6.21	Transmembrane	229 - 245 (223 - 247)
INTEGRAL	Likelihood	= -6.10	Transmembrane	44 - 60 (40 - 66)
INTEGRAL	Likelihood	= -4.51	Transmembrane	314 - 330 (312 - 331)

PERIPHERAL Likelihood = 0.90 212
modified ALOM score: 2.79

5

```

bacterial membrane --- Certainty=0.5585(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

10

ORF01027(379 - 1656 of 1968)
EGAD|33392|34706(15 - 423 of 427) cymF protein {Klebsiella oxytoca}
GP|854233|emb|CAA60005.1||X86014 cymF {Klebsiella oxytoca} PIR|S63615|S63615 malF protein
homolog cymF - Klebsiella oxytoca
%Match = 23.8
%Identity = 41.3 %Similarity = 64.5
Matches = 171 Mismatches = 140 Conservative Sub.s = 96

20

132 162 192 222 252 282 312 342
VLLFLAILTVVKSNLAILTVNVSNNISIKTSLKONSTRSRVMR*GEYSSFOLRVLPISYFVK*OLKETIMNKKLISLDGMSLK

ML

25

372 402 432 462 492 522 552 582
EVFQKGLD L A T K L S F L I M G L A N L K N K Q I V K G L L F L I S E I L F L I T F V Y Q V I P A V K G L I S L G T Q E Q G M T T K T V D G I K I Q V A T Q
 | | | | | | : : : | | : | : | : : : : | : | | : | | | | | |
L S E G K S M R I F P A S F L I M G A T Q L I S G H W I K G S V F L L F Q I - V V I S N I N L L L N A T Q G L I T L G T V A Q - - - - - T R S G F D I - - - V A
 20 30 40 50 60 70

30

612 642 672 702 732 762 792 822
GDN\$MLM\$LIFGLASLIFXCVPAYIYWSNIXSAHLLTLKEEGREIP\$FKDKIKSLTDGRFHM\$TMSIPLIGVLLFTILPL
|||:::||: |: ::|| :|| || : : | : :: : | || :: : | : | : :
GDNSIFMLVEGVVA\$IFLF\$SIFVYLNI-----KDAQVCEKCHQS\$FTEQLRLTYDNR\$FATIMLAPAFIACTA\$FIIMP\$

90 100 110 120 130 140

35

852 879 909 939 966 996 1026 1056
VYMICLAFTNYDH-NHLPPKSLFDWVGFAFNGNIFSGRMAS-TFFPVLWSWTLIWAVFATVINFFGGIILALLINTKGLKF
: : ::||| :||:| |||: || :| :| || : ||::|| |||: || :||| : :|
IITVLVSLTNSAPHHPKKNLVDWVGLKNFITLELRIWSKTFVIGIGVWTVLWFAFFATLCTCSFGFLIALALENKKIIA
160 170 180 190 200 210 220

40

1086 1116 1146 1176 1206 1236 1266 1296
KKMWRITFIVITMAVPQFISLLIMRNLLSDAGPVNALLIKWGLISSAHLPLFLSPVWAKFSIIFVNMWVGIPVTMLVATG
|| || :||: ||: ||: |||| ||: ||||: || || : ||||: || ||: ||: |||| ||: ||
KKAWRVVFILPYAIPAFVTLILFRLLNLGIGPVNSTLNSWG---IDSIGFLSDPLIAKMTVIADVSVWVGAPYFMLLITG
240 250 260 270 280 290 300

45

1326 1356 1386 1416 1446 1476 1506 1536
IIMNLPAEQIEAAEIDGANKFQVFSITFPQILLIMTPTLIQQFIGNNFNVIYLLTQGQPNTSTYYQAGSTDLLVTWL
: |:| : ||:|:||:|| | : ||:| : | : |:| : ||| |||:|:| | | || |:|:|:
AMTNIPRDLYEAESVDGASKFQFREITLPMVLHQVAPSLVMTFAHNFNNFGAIYLLTEGGPINPEYRFAGHTDILITIWI

320 330 340 350 360 370 380

50

1566 1596 1626 1656 1686 1716 1746 1776
 YNLTVTAAADYNLASVVGILIFILSAVFSLLAYTRTNSYKEGAAK**IRKNVLTLLLLFIFY**YYQLCGSFPLFGSFSQAS
 | ||: | :|||: |:||: :||: : | |:||
 YKLTLDFFQQYQIASVISIIIFLFLSIFAIWQFRMRKSKFEDVGM
 400 410 420

60

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1840-

Example 1641

A DNA sequence (GBSx1736) was identified in *S.agalactiae* <SEQ ID 5075> which encodes the amino acid sequence <SEQ ID 5076>. This protein is predicted to be maltose/maltodextrin-binding protein precursor. Analysis of this protein sequence reveals the following:

```

5   Possible site: 41
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -3.98    Transmembrane    25 - 41 ( 24 - 43)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.2593(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

15 A related GBS nucleic acid sequence <SEQ ID 9999> which encodes amino acid sequence <SEQ ID 10000> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA26925 GB:L08611 MalX [Streptococcus pneumoniae]
  Identities = 117/418 (27%), Positives = 186/418 (43%), Gaps = 43/418 (10%)

20  Query: 15  TKMEKNTWKKLLVSTAALSVVAGGAIAATHSNSVDAASKTTIKLWVPTDSKASYKAIVKK 74
      +K K+T V+ A+L +VA G+ A ++ + ++V K+ + + K
  Sbjct: 3    SKFMKSTAVLGTVTLASLLLVACGSKTADKPADSGSSEVKELTVYVDEGYKSYIEEVAKA 62

  Query: 75  FZKE-NKGVTVMKIESNDSKAQENVKKDPSKAADVFSPLPHDQLGQLVESGVIQEIPEQYS 133
      ++KE VT+K ++ + ++ DV P+D++G L G + E+ + S
  Sbjct: 63  YEKEAGVKVTLKTGDALGGLDKLSLDNQSGNVDPVMMAPYDRVGSGLSDGQLSEV--KLS 120

  Query: 134 KEIAKNDTKQSLTGAQYKKGKTYAFPFGIESQVLYYNKTKLTADDVKSYETITSKGKFGXQ 193
      +DT +SL A GK Y P IES V+YNNK L D K++ + + K
  Sbjct: 121 DGAKTDDTTKSLVTAA-NGKVYGAFAVIESLVMYYNKD-LVKDAPKTFADLENLAKDSKY 178

  Query: 194 LKAA-----NSYVTGPXFLSVGDTLFGKSGEDAKGTNMGNEAGVSVL----- 235
      A N Y T G +FG++G+DAK N+ ++ +
  Sbjct: 179 AFAGEDGKTTAFLADWTNIFYTYTGLLAGNGAYVFGQNGKDAKDIGHLANDGSIAGINYAKS 238

35  Query: 236 ---KWIADQKKNDFVNLTAENTMSKFGDGSVHAFESGPWDYDAAKKAVGEDKIGVAVYP 292
      KW + +G NL ++F +G A GPW A K A + GVA P
  Sbjct: 239 WYEKWPFGMDTEGAGNLI----QTQFQEGKTAAIIDGPWKAQAFKDA--KVNYGVATIP 292

40  Query: 293 TMKIGDKVEQQKAFGLGVKLYAVNQAPAGSNTKRISASYKLAAYLTNAESQKIQFEKRHIV 352
      T+ G + AF G K + + QA K + AS K +L E QK+ ++K + +
  Sbjct: 293 TLPNGK---EYAAFGGGKAWVIPQA-----VKNLEASQKFVDFLVATEQQKVLVDKTNIEI 344

  Query: 353 PANSSIQSSDSVQKDELAKAVIEMGSSDKYTTVMPKLSQMSTFWTESAAILSDTYSGK 410
      PAN+ +S + DEL AVI+ K T +P +SQMS W + +L D SG+
45  Sbjct: 345 PANTEARSYAEGKNDELTTAVIK---QFKNTQPLEPNISQMSAVWDPKNMLFDAVSGQ 399

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5077> which encodes the amino acid sequence <SEQ ID 5078>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 28

    >>> May be a lipoprotein

    ----- Final Results -----
55  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

-1841-

>GP:AAA26925 GB:L08611 MalX [Streptococcus pneumoniae]

Identities = 126/423 (29%), Positives = 191/423 (44%), Gaps = 50/423 (11%)

5 Query: 13 SLTLASTLLVCGSGSKDK--KEAGADSKTIKLWVPTGSKSYADTIK-FEKDSGYTVK 69
 ++TLAS LLV CGS + DK ++ K + ++V G KSY + +AK +EK++G V
 Sbjct: 14 TVTLASLLLACGSKTADKPADSGSSEVKELTVYVDEG-YKSYIEEVAKAYEKEAGVKVT 72

10 Query: 70 VVESEDPKAQEKIKKD--ASTAADVFSLPHDQLGQLVESGTIQEVPEKYNKEIAATSTDQ 127
 + + +K+ D + DV P+D++G L G + EV K + T +
 Sbjct: 73 LKTGDALGGLDKLSLDNQSGNVPDVM MAPYDRVSGSLGSDGQLSEV--KLSDGAKTDDTTK 130

15 Query: 128 ALVGAQYKGKTYAFPPGIESQVLFYNKSKLAAEDVTSYD---TITTKATFGGTFKQ--- 180
 +LV A GK Y P IES V++YNK + T D +K F G +
 Sbjct: 131 SLVTAA-NGKVYGAPAVIESLVMYYNKDLVKDAPKTFADLENLAKDSKYAFAGEDGKTTA 189

20 Query: 181 -----ANTYATGPLFMSVGNLTFGENGEDVKGTNWGNEKGAAVL-----KWIADQAS 227
 N Y T L G +FG+NG+D K N+ A + KW
 Sbjct: 190 FLADWTNIFYTYTGLLAGNGAYVFGQNGKDAKDIGLANDGSIAGINYAKSWYKWPKGMD 249

25 Query: 228 NKGFSVSLDANNVMSKFGDGSVASFESGPWDYEAQAIGKENLGVAIYPKVTIGGETVQ 287
 +G N + ++F +G A+ GPW +A + A K N GVA P + G E
 Sbjct: 250 TEG----AGNLIQTQFQEGKTAAILDGPWKAQAFKDA--KVNYGVATIPTLPNGKE---Y 300

30 Query: 288 KAFLGVKLYAVNQAPAKGDTKRIAASYKLASYLTNAESQENQFKTRNIVPANKEVQSSEA 347
 AF G K + + QA K + AS K +L E Q+ + N +PAN E +S
 Sbjct: 301 AAFGGGKAWVIPQA-----VKNLEASQKFVDFLVATEQQKVLVDKTNEIPANTEARSYAE 355

35 Query: 348 VQSNEAKTVITMGSSSDYTVVMPKLSQMGTFTWTESAAILSDAFNG----KIKENDYLTK 403
 +++EL VI + T +P +SQM W + +L DA +G K ND +T
 Sbjct: 356 GKNDLTTAVIKQFKN---TQPLPNISQMSAVWDPKNMFLDAVSGQKDAKTAANDAVTL 412

Query: 404 LQQ 406
 +++
 Sbjct: 413 IKE 415

An alignment of the GAS and GBS proteins is shown below.

Identities = 278/415 (66%), Positives = 334/415 (79%), Gaps = 6/415 (1%)

40 Query: 21 TWKKLLVSTAALSVVAGGAIATHSNSVD---AASKTTIKLWVPTDSKASYKAIVKKFZ 76
 +W+K++V A+L++ A + S S D A TIKLWVPT SK SY + KF+
 Sbjct: 3 SWQKVIIVGGASLTLLASTLLVCGSGSKDKKEAGADSKTIKLWVPTGSKSYADTTAKFE 61

45 Query: 77 KENKGVTVKMIESNDSKAQENVVKDPSKAADVFSLPHDQLGQLVESGVIQEIPEQYSKEI 136
 K++ G TVK++ES D KAQE +KKD S AADVFSLPHDQLGQLVESG IQE+PE+Y+KEI
 Sbjct: 62 KDS-GYTVKVVESEDPKAQEKIKKDASTAADVFSLPHDQLGQLVESGTIQEVPEKYNKEI 120

50 Query: 137 AKNDTKQSLTGAQYKGKTYAFPPGIESQVLYYNKTKLTADDVKS YETITTSKGKFGXQLKA 196
 A T Q+L GAQYKGKTYAFPPGIESQVL+YNK+KL A+DV SY+TIT+K FG K
 Sbjct: 121 AATSTDQALVGAQYKGKTYAFPPGIESQVLFYNKSKLAAEDVTSYDTITTKATFGGTFKQ 180

55 Query: 197 ANSYVTGPXFLSVGDTLFGKSGEDAKGTNWGNEAGVSVLKWADQKKNDGFVNLTAEENTM 256
 AN+Y TGP F+SVG+TLFG++GED KGTNWGNE G +VLKWADQ N GFV+L A N M
 Sbjct: 181 ANTYATGPLFMSVGNLTFGENGEDVKGTNWGNEKGAAVLKWADQASNKGFSVSLDANNVM 240

60 Query: 257 SKFGDGSVHAFESGPWDYDAAKAVGEDKIGVAVYPTMKIGDKEVQQKAFLGVKLYAVNQ 316
 SKFGDGSV +FESGPWDY+AA+KA+G++ +GVA+YP + IG + VQQAFLGVKLYAVNQ
 Sbjct: 241 SKFGDGSVASFESGPWDYEAQAIGKENLGVAIYPKVTIGGETVQQAFLGVKLYAVNQ 300

65 Query: 317 APAGSNTKRISASYKLAAYLTNAESQKIQFEKRHIVPANSSIQSSDSVQKDELAKAVIEM 376
 APA +TKRI+ASYKLA+YLINAESQ+ QF+ R+IVPAN +QSS++VQ +ELAK VI M
 Sbjct: 301 APAKGDTRKRIAASYKLASYLTNAESQENQFKTRNIVPANKEVQSSEAVQSNEAKTVITM 360

Query: 377 GSSDKYTTVMPKLSQMSTFTWTESAAILSDTYSGKIKSSDYLRKLKQFDKDIATK 431
 GSS YT VMPKLSQM TFWTESAAILSD ++GKIK +DYL +L+QFDKDIA TK
 Sbjct: 361 GSSSDYTVVMPKLSQMGTFTWTESAAILSDAFNGKIKENDYLTFLKQFDKDIATK 415

-1842-

SEQ ID 5076 (GBS649) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 132 (lane 2 & 3; MW 76kDa) and in Figure 186 (lane 7; MW 76kDa).. It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 132 (lane 7; MW 51kDa) and in Figure 178 (lane 8; MW 51kDa).

- 5 GBS649-His was purified as shown in Figure 229, lane 8. Purified GBS649-GST is shown in Figure 245, lanes 6 & 73.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1642

- 10 A DNA sequence (GBSx1737) was identified in *S.agalactiae* <SEQ ID 5079> which encodes the amino acid sequence <SEQ ID 5080>. Analysis of this protein sequence reveals the following:

Possible site: 51
>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2462(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD02112 GB:AF039082 putative maltose operon transcriptional
repressor [Lactococcus lactis]
Identities = 43/61 (70%), Positives = 49/61 (79%)

- 25 Query: 2 VTIKDVAAGVNPSTVSRVLKDNASISSKTKERVKKAMEELGYVPNVAAQMLASGLTQN 61
VTIKDVA KAGVN STVSRV+KD++ IS KTK +V+KAM ELGY N AAQ+LASG T
Sbjct: 3 VTIKDVAAGVNPSTVSRVLKDNASISSKTKERVKKAMEELGYVPNVAAQMLASGLTNT 62

- Query: 62 I 62
I
30 Sbjct: 63 I 63

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5081> which encodes the amino acid sequence <SEQ ID 5082>. Analysis of this protein sequence reveals the following:

- 35 Possible site: 44
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -3.93 Transmembrane 269 - 285 (266 - 287)
----- Final Results -----
40 bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- 45 Identities = 53/62 (85%), Positives = 57/62 (91%)
Query: 1 MVTIKDVAAGVNPSTVSRVLKDNASISSKTKERVKKAMEELGYVPNVAAQMLASGLTQ 60
MVTIKDVA KAGVNPSTVSRVLKDN SIS KTKK+V+KAM +LGYVPNVAAQ+LASGLT
Sbjct: 26 MVTIKDVAAGVNPSTVSRVLKDNRSISMKTKEKVRKAMADLGYVPNVAAQILASGLTH 85
50 Query: 61 NI 62
NI
Sbjct: 86 NI 87

-1843-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1643

A DNA sequence (GBSx1738) was identified in *S.agalactiae* <SEQ ID 5083> which encodes the amino acid sequence <SEQ ID 5084>. Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -7.70	Transmembrane	14 - 30 (8 - 34)
INTEGRAL	Likelihood = -6.90	Transmembrane	66 - 82 (63 - 85)
INTEGRAL	Likelihood = -6.69	Transmembrane	110 - 126 (105 - 128)
INTEGRAL	Likelihood = -3.93	Transmembrane	132 - 148 (129 - 149)

----- Final Results -----

bacterial membrane	---	Certainty=0.4079(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9443> which encodes amino acid sequence <SEQ ID 9444> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC67260 GB:AF017113 Yvja [Bacillus subtilis]

Identities = 83/227 (36%), Positives = 140/227 (61%)

Query: 9 FGWDSAFFIMIINIPLLLLCYFGLGKQTFLLKTVYGSWIFPVFIKLTQSVPTLTHNPLLA 68
+G+++A+ IINIPL + LG + LKT+ GS P+ + LT+ + TH+ LLAA
Sbjct: 52 YGFEEAAYVQWIINIPLFIAGVILLGGKFLKTLGASVFLPLVVFLTRDIQPATHHELLAA 111

Query: 69 LFGGVIVGCGLGIVFWSDSSTGGTGIIIIQFLGKYTPISLGQGVILIDGLVTIVGFLAFDS 128
+FGGV +G G+GIV+ STGGT + Q + KY+ +SLG+ + +IDG++ + + F+
Sbjct: 112 IFGGVGIGIGIGIVYLKSGSTGGTALAAQIIHKYSGLSLGKCLAIIDGMIVVTAMIVFNI 171

Query: 129 DTFVMSIIGLITISYIINAIQTGFTTLSTVLIVSQEHQKIKTYINTVADRGVTEIPVKGG 188
+ +++++G+ S I+ +Q GF LI++++ Q +K + DRGVT+I GG
Sbjct: 172 EQGLYAMLGVYVSSKTIDVVQVGFNRSKMALIITKQEQAVKEAVLQKIDRGVTKISAVGG 231

Query: 189 YSGTNQIMLMTTIAGYEFKQLQEAIAEIDETAFITVTPTSQASGRGF 235
Y+ ++ +LM + EF KL++ + +IDE+AF+ V S+ G GF
Sbjct: 232 YTDDDRPILMCVVGQTEFTKLKQIVKQIDESAFVIVADASEVLGEGF 278

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5085> which encodes the amino acid sequence <SEQ ID 5086>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -6.21	Transmembrane	104 - 120 (101 - 123)
INTEGRAL	Likelihood = -3.93	Transmembrane	147 - 163 (142 - 167)
INTEGRAL	Likelihood = -3.29	Transmembrane	169 - 185 (169 - 186)

----- Final Results -----

bacterial membrane	---	Certainty=0.3484(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

>GP:AAC67260 GB:AF017113 Yvja [Bacillus subtilis]

Identities = 106/267 (39%), Positives = 169/267 (62%), Gaps = 1/267 (0%)

Query: 7 DLLLVLTIGSFITAIGFNTMFVDNHIASGGMVGIADVIVKALFGISPSLFLMASNIPLLLMC 66

-1844-

D + + IG+ ITA+ FN + N IA+GG+ GI+ ++++ +G + NIPL +
 Sbjct: 13 DYVYILIGAAITAVSFNVFLLPNKIAAGGVSGISTILQS-YGFEAAVQWIINIPLFIAG 71
 Query: 67 YFFLGKQNFIKTLYGWSWIYPIAIRSTNSLPTLTHNQLLAAIFGGIICGIGLGMVFWGNSS 126
 LG + +KTL GS P+ + T + TH++LLAAIFGG+ GIG+G+V+ G S
 Sbjct: 72 VILLGKGFGLKTLGASVFLPLVVFITRDIQPATHHELLAAIFGGVGIGIGIGIVLKGKS 131
 Query: 127 TGGTGILTQILHKYSPLSLGVAMTIVDGISVLMGFIALSADDVMYSTIGLFVIGYVISVM 186
 TGGT + QI+HKYS LSLG + I+DG+ V+ I + + +Y+ +G++V I V+
 Sbjct: 132 TGGTALAAQIIHKYSGLSLKGCLAIIDGMIVVTAMIVFNIEQGLYAMLGVYVSSKTIDVV 191
 Query: 187 ENGFDSSKNVMIISKDYQAIREYITTVMDRGVTKLPIRGYTTSDKIMLMAIVSSHELPT 246
 + GF+ SK +II+K QA++E + +DRGVTK+ GGYT D+ +LM +V E
 Sbjct: 192 QVGFNRSKMALIITKQEQAVKEAVLQKIDRGVTKISAVGGYTDDDRPILMCVVGQTEFTK 251
 Query: 247 LQEKILEIDDTAFIVVMPAAQVMGRGF 273
 L++ + +ID++AF++V A++V+G GF
 Sbjct: 252 LKQIVKQIDESAFVIVADASEVLGEF 278

An alignment of the GAS and GBS proteins is shown below.

Identities = 135/252 (53%), Positives = 190/252 (74%)

Query: 1 MAVSFHEVFGWDSAFFIMIINIPLLLL+CYFGLGKQTFKLT+VYGSWIFPVFIKLTQSVPTL 60
 +AV +FG + F+M NIPLLL+CYF LGKQ F+KT+YGSWI+P+ I+ T S+PTL
 Sbjct: 39 IAVVIKALFGISPSLFLMASNIPLLLMCYFPLGKQNFIKTLYGWSWIYPIAIRSTNSLPTL 98
 Query: 61 THNPLLAALFGGVIVGCGLGIVFWSDSSTGGTGIIIQFLGKYTPISLGQGVILIDGLVTI 120
 THN LLAA+FGG+I G GLG+VFW +SSTGGTGI+ Q L KY+P+SLG + ++DG+ +
 Sbjct: 99 THNQLLAAIFGGIICGIGLGMVFWGNSSSTGGTGILTQILHKYSPLSLGVAMTIVDGISVL 158
 Query: 121 VGFLAFSDTVMFSIIGLITISYIINAIQTGFTTLSTVLIVSQEHQKIKTYINTVADRGV 180
 +GF+A +D VM+S IGL I Y+I+ ++ GF + V+I+S+++Q I+ YI TV DRGV
 Sbjct: 159 MGFIALSADDVMYSTIGLFVIGYVISVMENGFDSSKNVMIISKDYQAIREYITTVMDRGV 218
 Query: 181 TEIPVKGSGYSGTNQIMLMTTIAGYEFAKLQEAIAEIDETAFTVTPTSQASGRGFSLQKN 240
 T++P++GGY+ +++IMLM ++ +E LQE I EID+TAFI V P +Q GRGFSL K
 Sbjct: 219 TKLPIRGYTTSDKIMLMAIVSSHELPTLQEKILEIDDTAFIVVMPAAQVMGRGFSLTKQ 278
 Query: 241 HGRLEDEDILMPM 252
 + R D+D+L+PM
 Sbjct: 279 YKREDKDVLLPM 290

A related GBS gene <SEQ ID 8871> and protein <SEQ ID 8872> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6
 McG: Discrim Score: 1.57
 GvH: Signal Score (-7.5): -2.56
 Possible site: 56
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 4 value: -7.70 threshold: 0.0
 INTEGRAL Likelihood = -7.70 Transmembrane 14 - 30 (8 - 34)
 INTEGRAL Likelihood = -6.90 Transmembrane 66 - 82 (63 - 85)
 INTEGRAL Likelihood = -6.69 Transmembrane 110 - 126 (105 - 128)
 INTEGRAL Likelihood = -3.93 Transmembrane 132 - 148 (129 - 149)
 PERIPHERAL Likelihood = 3.71 37
 modified ALOM score: 2.04
 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-1845-

[illegible]

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1644

40 A DNA sequence (GBSx1739) was identified in *S.agalactiae* <SEQ ID 5087> which encodes the amino acid sequence <SEQ ID 5088>. This protein is predicted to be ABC transporter, ATP-binding protein (b0820). Analysis of this protein sequence reveals the following:

```

Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3122(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC24918 GB:AF012285 YkpA [Bacillus subtilis]
Identities = 355/540 (65%), Positives = 451/540 (82%), Gaps = 4/540 (0%)

55 Query: 1 MLTVSDVSLRFSDRKLFDDEVNINFAGNTYGLIGANGAGKSTFLKILAGDIEPTTGHIAL 60
M+ V++VSLRF+DRKLF++VNI FT GN YGLIGANGAGKSTFLK+L+G+IEP TG + +
Sbjct: 1 MIAVNNVSLRFADRKLFDDEVNIKFTPGNCYGLIGANGAGKSTFLKVLSGEIEPQTGDVHM 60

60 Query: 61 GPDERLSVLRQNHFDYEDERVIDVVMGNETLYSIMKEKDAIYMKEDFSDEGDGVRAAELE 120
P ERL+VL+QNH+YE+ V+ VVMG++ LY +M+EKDAIYMK DFSDEGD+RAAELE
Sbjct: 61 SPGERLAVLQNHFEYEEYEVLLKVVIMGHKRLYEVMQEKDAIYMKPDFSDEGDGIRAAELE 120

-1846-

Query: 121 GEFAELGGWEAESEASQLLQNLNISEELHYQNMSSELANGDKVKVLLAKALFGKPDVLLLD 180
 GEFAEL GWEAESEA+ LL+ L ISE+LH + M++L +KVKVLLA+ALFGKPDVLLLD
 Sbjct: 121 GEFAELNGWEAESEAAILLKGLGISEDLHTKKMADLGGSEKVKVLLAQALFGKPDVLLLD 180

Query: 181 EPTNGLDIQSITWLEDFLIDFENTVIVVSHDRHFLNKVCTHMAADLDFGKIKLFGVGNDFW 240
 EPTN LD+Q+I WLE+FLI+FENTVIVVSHDRHFLNKVCTH+ADLDF KI+++VGNYDFW
 Sbjct: 181 EPTNHLDLQAIQWLEEFLLNFENTVIVVSHDRHFLNKVCTH+ADLDFNKI+QIYVGNDFW 240

Query: 241 KESSELAARLQADRNAKAEKIKQLQEFVARFSANASKSQATSRRKKMLDKIELEEIVPS 300
 ESS+LA +L + N K EE+IKQLQEFVARFSANASKSQATSRRKK+L+KI L++I PS
 Sbjct: 241 YESSQLALKLSQEANKKKEEQIKQLQEFVARFSANASKSQATSRRKKLEKITLDDIKPS 300

Query: 301 SRKYPFVNFKAEREMGNDLLTVENLSVTIDGEKILDNISFILRPGDKTALIGQNDIQTTA 360
 SR+YP+VNF ERE+GND+L VE L+ TIDG K+LDN+SFI+ DK A G+N++ T
 Sbjct: 301 SRRYPYVNFPTPEREIGNDVLVRVEGLTKTIDGVKVLNVSFIMNREDKIAFTGRNELAVTT 360

Query: 361 LIRALMGDIEYE-GTIKWGVTTSSRSLPKDNRDFASGE-SILEWLRQFASKEEDDNTFL 418
 L + + G++E + GT KGWVTT++Y PKDNS F + +++WLRQ+ S + +FL
 Sbjct: 361 LFKIISGEMEADSGTFKWGVTTSQAYFPKDNSEYFEGSDNLNVDWLRQY-SPHDQSESFL 419

Query: 419 RGFLGRMLFSGDEVNKSVMNLVSGGEKVRVMSKMLMLLKSNNVLVDDPTNHLDLSEISSIN 478
 RGFLGRMLFSG+EV+K NVLSGGEKVR ML SK ML +N+L+LD+PTNHLDLSEI++LN
 Sbjct: 420 RGFLGRMLFSGEEVHKANVLVSGGEKVRMLSKAMLSGANILILDEPTNHLDLSEITALN 479

Query: 479 DGLKDFKESIIIFASHDHEFIQTLANHHIIVLSKNGVIDRIDETIDEFLENTEVQAKVAQLW 538
 +GL FK +++F SHDH+F+QT+AN II ++ NG++D+ +YDEFLEN +VQ K+ +L+
 Sbjct: 480 NGLISFKGAMLFTHSHDHQFVQTIANRIIEITPNGIVDK-QMSYDEFLENADVQKKLTLEY 538

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5089> which encodes the amino acid sequence <SEQ ID 5090>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3124(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 497/539 (92%), Positives = 525/539 (97%)

Query: 1 MLTVSDVSLRFSRKLFDENVINFTAGNTYGLIGANGAGKSTFLKILAGDIEPTGHIAL 60
 +LTVSDVSLRFSRKLFD+VNI FTAGNTYGLIGANGAGKSTFLKILAGDIEP+TGHI+L
 Sbjct: 1 LLTVSDVSLRFSRKLFDENVINFTAGNTYGLIGANGAGKSTFLKILAGDIEPTGHISL 60

Query: 61 GPDERLSVLRQNHFDYEDERVIDVVMGNETLYSIMKEKDAIYMKEDFSDEGVRAAELE 120
 GPDERLSVLRQNHFDYE+ER IDVVMGNE LY+IMKEKDAIYMK DFS+EDGVRAAELE
 Sbjct: 61 GPDERLSVLRQNHFDYEEERAIDVVMGNEQLYNIMKEKDAIYMKADFSEEDGVRAAELE 120

Query: 121 GEFAELGGWEAESEASQLLQNLNISEELHYQNMSSELANGDKVKVLLAKALFGKPDVLLLD 180
 G FAELGGWEAESEASQLLQNLNI E+LHYQNMSSELANGDKVKVLLAKALFGKPDVLLLD
 Sbjct: 121 GIFAELGGWEAESEASQLLQNLNIPEDLHYQNMSSELANGDKVKVLLAKALFGKPDVLLLD 180

Query: 181 EPTNGLDIQSITWLEDFLIDFENTVIVVSHDRHFLNKVCTHMAADLDFGKIKLFGVGNDFW 240
 EPTNGLDIQSI+WLEDFLIDFENTVIVVSHDRHFLNKVCTHMAADLDFGKIKLFGVGNDFW
 Sbjct: 181 EPTNGLDIQSIWLEDFLIDFENTVIVVSHDRHFLNKVCTHMAADLDFGKIKLFGVGNDFW 240

Query: 241 KESSELAARLQADRNAKAEKIKQLQEFVARFSANASKSQATSRRKKMLDKIELEEIVPS 300
 K+SSELAARLQADRNAKAEKIK+LQEFVARFSANASKSQATSRRKKMLDKIELEEIVPS
 Sbjct: 241 KQSELAARLQADRNAKAEKIKELQEFVARFSANASKSQATSRRKKMLDKIELEEIVPS 300

Query: 301 SRKYPFVNFKAEREMGNDLLTVENLSVTIDGEKILDNISFILRPGDKTALIGQNDIQTTA 360
 SRKYPF+NFKAEREMGND LTVENLSVTIDGEKI+DNISFILRPGDK A+IGQNDIQTTA
 Sbjct: 301 SRKYPFINFKAEREMGNDLTVENLSVTIDGEKIDNISFILRPGDKAIIIGQNDIQTTA 360

-1847-

Query: 361 LIRALMGDIEYEGTIKGVTTTSRSLPKDNSRDFASGESILEWLRQFASKEEDNTFLRG 420
 L+RAL DI+YEGTIKGVTTTSRSLPKDNS+DFA+ ESILEWLRQFASK EDD+TFLRG
 Sbjet: 361 LMRALADDIDYEGTIKGVTTTSRSLPKDNSKDFATEESILEWLRQFASKGEDDDTFLRG 420

Query: 421 FLGRMLFSGDEVNKSNNVLSGGGEKVRVMSKLMMLKSNVLVLDPTNHLDLSEISSLNDG 480
 FLGRMLFSGDEV KSNVNLSSGGGEKVRVMSKLMMLKSNVL+LDDPTNHLDLSEISSLNDG
 Sbjet: 421 FLGRMLFSGDEVKSNVNLSSGGGEKVRVMSKLMMLKSNVLILDDPTNHLDLSEISSLNDG 480

Query: 481 LKDFKESIIIFASHDHEFIQTLANHIIIVLSKNGVIDRIDETYDEFLENTEVQAKVAQLWK 539
 +KDFKES+IFASHDHEFIQT+ANHI+V+SKNGVIDRIDETYDEFI+N EVQA+VA+LWK
 Sbjet: 481 IKDFKESVIFASHDHEFIQTIANHIVVISKNGVIDRIDETYDEFIDNPEVQARVAELWK 539

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1645

A DNA sequence (GBSx1740) was identified in *S.agalactiae* <SEQ ID 5091> which encodes the amino acid sequence <SEQ ID 5092>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.71	Transmembrane	14 - 30 (8 - 35)
INTEGRAL	Likelihood = -7.70	Transmembrane	384 - 400 (382 - 403)
INTEGRAL	Likelihood = -7.22	Transmembrane	412 - 428 (408 - 433)
INTEGRAL	Likelihood = -5.73	Transmembrane	163 - 179 (155 - 180)
INTEGRAL	Likelihood = -5.52	Transmembrane	322 - 338 (320 - 344)
INTEGRAL	Likelihood = -5.10	Transmembrane	297 - 313 (290 - 314)
INTEGRAL	Likelihood = -4.41	Transmembrane	360 - 376 (357 - 377)
INTEGRAL	Likelihood = -4.35	Transmembrane	438 - 454 (437 - 455)
INTEGRAL	Likelihood = -4.09	Transmembrane	136 - 152 (136 - 153)
INTEGRAL	Likelihood = -3.35	Transmembrane	110 - 126 (106 - 128)
INTEGRAL	Likelihood = -2.28	Transmembrane	232 - 248 (232 - 248)
INTEGRAL	Likelihood = -1.81	Transmembrane	832 - 848 (832 - 848)
INTEGRAL	Likelihood = -1.12	Transmembrane	200 - 216 (200 - 216)

----- Final Results -----

bacterial membrane --- Certainty=0.4885(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC14608 GB:U95840 transmembrane protein Tmp5 [Lactococcus lactis]

Identities = 140/260 (53%), Positives = 182/260 (69%), Gaps = 6/260 (2%)

Query: 16 SFLLPFIIIVCILFTKNIYWGSPPTILASDGFHQYVIFNQALRNILH--GSNSLFYTFSTS 73
 SF +P I++V +L IYWGS +ILA D +HQYV + RNILH GS YTFSTS
 Sbjet: 14 SFFIPLIILMVIVLAMTGIYWGSSRSILAGDAYHQYVAIHSLYRNILHSGSGQGFYTFSTS 73

Query: 74 GLGLNFYALSSYYLGSFLSPIVYFFNLKNMPDAIYLLTICKIGLIGLSMFVTLCKRHCKV 133
 GLGLN YA S+YY+GSFL P +FF++K+MPDA+YL TI K GLIGLS FV+ + K+
 Sbjet: 74 GLGLNLYAFSAYYMGSLMPFTFFFDVKSMPDALYLFITIKFGLIGLSFVSFKNMYQKL 133

Query: 134 NRVLLLVISTCYLSMSFSISQIEINMWLDVFILIPVLVGVLDQLLWERKPIYFLSLTAL 193
 + + +L IST ++LMSF SQ+EI MWLDVFIL+PL++ G+ +L+ ERK LYF+SL L
 Sbjet: 134 SNLTVLSISTAFALMSFLTSQLEITMWLDVFILLPLIHWGLHRLMDERKRWLYFVSLIIL 193

Query: 194 FIQNYFYFGFMTAIFTSLYFIVQITRNTDSKVAFKQLHFTFLSLLAGMTSSIMILPTYFD 253
 FIQNYFYFGFM AIF LYF + R T K ++ + L F S LAG+ S IM+LP Y D
 Sbjet: 194 FIQNYFYFGFMAIFLVLYF---LARMYKWSWTKVLDVVSSTLAGIASLIMLLPMYLD 250

Query: 254 L-TTHGEKLTQVSKMFTENS 272
 L + + + L+ +S +FTENS

-1848-

Sbjct: 251 LKSNNSDALSTLSGIFTENS 270

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5093> which encodes the amino acid sequence <SEQ ID 5094>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 51
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -9.29    Transmembrane    15 - 31 ( 6 - 35)
      INTEGRAL    Likelihood = -8.81    Transmembrane    201 - 217 ( 196 - 220)
      INTEGRAL    Likelihood = -6.79    Transmembrane    410 - 426 ( 402 - 428)
10     INTEGRAL    Likelihood = -6.05    Transmembrane    230 - 246 ( 227 - 252)
      INTEGRAL    Likelihood = -5.57    Transmembrane    161 - 177 ( 153 - 178)
      INTEGRAL    Likelihood = -4.46    Transmembrane    291 - 307 ( 290 - 311)
      INTEGRAL    Likelihood = -3.82    Transmembrane    133 - 149 ( 130 - 151)
      INTEGRAL    Likelihood = -3.77    Transmembrane    380 - 396 ( 376 - 400)
15     INTEGRAL    Likelihood = -3.61    Transmembrane    105 - 121 ( 103 - 124)
      INTEGRAL    Likelihood = -3.45    Transmembrane    832 - 848 ( 830 - 848)
      INTEGRAL    Likelihood = -2.66    Transmembrane    436 - 452 ( 435 - 453)
      INTEGRAL    Likelihood = -2.13    Transmembrane    318 - 334 ( 314 - 336)
      INTEGRAL    Likelihood = -1.54    Transmembrane    356 - 372 ( 355 - 372)
20     INTEGRAL    Likelihood = -0.27    Transmembrane    80 - 96 ( 80 - 96)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAC14608 GB:U95840 transmembrane protein Tmp5 [Lactococcus lactis]
Identities = 134/269 (49%), Positives = 183/269 (67%), Gaps = 8/269 (2%)
30
Query: 5  NKWIIAGLASFLFPLSIIFIILLSMGIYYNSDKTILASDAFHQYVIFAQNFNIMH--GS 62
      NKW + LASF PL ++ I+L GIY+ S ++ILA DA+HQYV +RNI+H GS
Sbjct: 7  NKWAL--LASFFIPLILMVIVLAMTGIYWGSSRSILAGDAYHQYVAIHSLYRNILHSGGS 64

Query: 63 DSFFYTFTSGLGINFYALMCYYLGSSFFSPLLFFFNLTSMPDAIYLFSLIKFGLIGLAACY 122
      F YFTFTSGLG+N YA YY+GSF P FFF++ SMPDA+YLF+IKFGLIGL++
Sbjct: 65 QGFLYTFTSGLGLNLYAFSAIYMGSLMPFTFFFDVKSMPDALYLFSLIKFGLIGLSSFV 124

Query: 123 SFHRLYPKISAFMLISISVFYSLMSFLTSQMELNSWLDVFILLPLVILGLNKLITENKTR 182
      SF +Y K+S ++SIS ++LMSFLTSQ+E+ WLDVFILLPL+I GL++L+ E K
40     Sbjct: 125 SFKNMYQKLSNLTVLSISTAFALMSFLTSQLEITMWLDVFILLPLIHWGLHRLMDERKRW 184

Query: 183 TYVLSISLLFIQNYIFGYMIALFCILYALVCLRLNDFNKMFIQFVFRFTAVSICAALTS 242
      Y++S+ +LFIQNYIFG+M+A+F +LY L R+ + + F S A + S
45     Sbjct: 185 LYFVSLILFIQNYIFGFMVAIFLVLYFLA---RMTYEKWSWTKVLDVVSSTLAGIASL 241

Query: 243 LVILPTYLDL-STYGENLSPIKQLVINNA 270
      +++LP YLDL S + LS + + T N+
50     Sbjct: 242 IMLLPMYLDLKSNNSDALSTLSGIFTENS 270

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 432/836 (51%), Positives = 569/836 (67%), Gaps = 2/836 (0%)
55
Query: 16 SFLLPFIIIVCILFTKNIIYWGSPPTTILASDGFHQYVIFNQALRNILHGSNSLFYTFTSGL 75
      SFL P II IL + IY+ S TILASD FHQYVIF Q RNI+HGS+S FYTFTSGL
Sbjct: 14 SFLFPLSIIFIILLSMGIYYNSDKTILASDAFHQYVIFAQNFNIMHGSDSFFYTFTSGL 73

Query: 76 GLNFYALSSYYLGSLSPIVYFFNLKNMPDAIYLLTICKIGLIGLSMFVTLCKRHCKVNR 135
      G+NFYAL YYLGSF SP+++FFNL +MPDAIYL T+ K GLIGL+ + + + K++
60     Sbjct: 74 GINFYALMCYYLGSSFFSPLLFFFNLTSMPDAIYLFSLIKFGLIGLAACYSFHRLYPKISA 133

Query: 136 VLLLIVISTCYSLMSFSISQIEINMWLDVFILIPLVVLGVDQLWERKPILYFLSLTALFI 195
      L++ IS YSLMSF SQ+E+N WLDVFIL+PLV+LG+++L+ E K Y+LS++ LFI
Sbjct: 134 FLIMISISVFYSLMSFLTSQMELNSWLDVFILLPLVILGLNKLITENKTRTYVLSISLLFI 193

```

-1849-

Query: 196 QNYYFGFMTAIFTSLYFIVQITRNTDSKVAFKQLHFTFLSLLAGMTSSIMILPTYFDLT 255
 QNYYFG+M A+F LY +V + R D F F+ FT +S+ A +TS+++ILPTY DL+
 Sbjct: 194 QNYYFGYMIALFCILYALVCLLRNDFNMKMFIAFVRFTAVSICAALTSALVILPTYLDLS 253

5
 Query: 256 THGEKLTQVSKMFTENSWYMDLFAKNMIGAYDTTKFGSIPMIYVGLPLLLSLLYFTIKE 315
 T+GE L+ + ++ T N+W++D+ AK IG YDTTKF ++PMIYVGL PL+LS++YFT++
 Sbjct: 254 TYGENLSPIKQLVTNNAWFLDIPAKLSIGVYDTTKFNALPMIYVGLFPLMLSIVYFTLES 313

10
 Query: 316 VPRRTRLAYGFLIIFVIASFYITPLDLFWQGMHAPNMFLHRYSWVLSVLICLLAAECLEY 375
 +P + +LA L+ F+I SFY+ PLDLFWQGMH+PNMFLHRY+W S++I LLA E L
 Sbjct: 314 IPLKIKLANACLLTFIISFYQLPLDLFWQGMHSPNMFLHRYAWSFSIVILLACETLSR 373

15
 Query: 376 LDNISWKKILGVNLILVSGFIITFLFKKHYYHLNLELLLTTLTFLSAYIILTISFVSKQI 435
 L ++ K + L+ + + F + Y++L L L LL++ L Y I SF + QI
 Sbjct: 374 LKEVTQIKAGFAFIFLIILTSLPYSFSQQYNFLPLTLFLLSVFLLGYTISLFSFRNSQI 433

20
 Query: 436 PKLVFPYFLIGFVLEMTLNTFYQLNSLNDEWIFPSRQYAKYNHSISKLVKRTERNNST 495
 P F++ F +LE LNT+YQL +N EW FPSRQ Y I+ LV +N+
 Sbjct: 434 PSTFISAFILIFSLESGLNTYYQLQGINKKEWGFPSRQIYNSQLKDINNLVNSVSKNSQP 493

25
 Query: 496 FFRTERWLGGTGNDSMKYNYNGISQFSSIRNRSSSQVLDRLGFKSDGTNLNLRYQNNTLI 555
 FFR ER L QTGNDSMK+NY GISQFSS+RNR SS +LDRLGF+S GTNLNLRYQNNT+I
 Sbjct: 494 FFRMERLLPQTGNDSMKFNYYGISQFSSVRNRLSSSLDLRLGFKSGTNLNLRYQNNTII 553

30
 Query: 556 ADSLFGVKYNLTEYPFDKFGFIKKAQDKQTILYKNQFASQLAILTNQVYQDKPFTVNTLD 615
 DSL G+KYNL+E P +KFGF K T LY+N ++S LAILT VY+D VNTLD
 Sbjct: 554 MDSLLGIKYNLSEGPPNKFQFTKLKTSQNTTLYQNHYSPLAILTRNVYKDVNLNVNTLD 613

35
 Query: 616 NQTTLNLQSLGKETYFEHLIPNSVSGQTTLNKQVFK-KNKQGNTEITYNITIPKNSQL 674
 NQT LLNQLSG TYF +SG N Q+ + + Q + + Y I IPK+SQL
 Sbjct: 614 NQTKLLNLQSLGKSLTYFNLPQAQLISGANQFNGQISAQASDYQNSVTLNYQINIPKHSQL 673

40
 Query: 675 YVSMPPFINFNEENKIVQISVNNGPFPVNTLDNAYSFFNIGSFAENSRIKVKQFPHNDQ 734
 YVS+P I F+N + K ++I +N F+ T DNAYSFF++G FA+ F FP N Q
 Sbjct: 674 YVSIPIIIFSNPDAKEMRIQTDNHNFI-YTTDNAYSFFDLGYFADAKVATFSFVFPKNKQ 732

45
 Query: 735 VSFPIPHFYGLKLEAYQKAMTVINKRKVKVRTDHNVIANYSPPNRSSLFFFTIPYDRGWK 794
 +SF PHFY L +E+Y +AM I ++ V N VI +Y S + SL FT+PYD+GW
 Sbjct: 733 ISFKEPHFYSLSIESYLEAMNSIKQKNVHTYAKSNTVITDYNSTKGSLLIFTLPYDKGWS 792

Query: 795 AYQNKEIKIFKAQKGFMKINIPKGGKVTILIFIPYGFKFGVGLSITGIVLFTVYY 850
 A ++ K + + KAQ GF+ + IPKGGK+V L FIP GFK G+ LS GI+ + + Y
 Sbjct: 793 AQKDGKNLPVKAQGGFLSVTIPKGGKRVILTIFIPNGFKLGLSLSCVGIIAYMLLY 848

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1646

A DNA sequence (GBSx1741) was identified in *S.agalactiae* <SEQ ID 5095> which encodes the amino acid sequence <SEQ ID 5096>. Analysis of this protein sequence reveals the following:

Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4624 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < .succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45340 GB:AF000658 ORF1 [Streptococcus pneumoniae]
 Identities = 111/159 (69%), Positives = 136/159 (84%)

-1850-

Query: 1 MKLKIIITVGKLKEKYLKEGVAEYQKRLNRFSKIETIELADEKTPDKASISENQRILDIEG 60
 MK+K++TVGKLKEKYLK+G+AEY KR++RF+K E IEL+DEKTPDKAS SENQ+IL+IEG
 Sbjct: 1 MKIKVVTVGKLKEKYLKDGIAEYSKRISRFAKFEMIELSDEKTPDKASESENQKILEIEG 60

5 Query: 61 ERILSKIGERDYGVLAEIGKQLPSESFSHLIDQKMISGYSTITTFVIGGSLGLSQQVKKR 120
 +RILSKI +RD+VI LAIEGK SE FS +++ I G+ST+TF+IGGSLGLS VK R
 Sbjct: 61 QRILSKIADRDVFIVLAIEGKTFSEEFQKLEETSIGKFSTLTFIIGGSLGLSSSVKNR 120

10 Query: 121 ADYLMSFGLLTLPHQLMKLVLMQIYRAFMRQGTPTYHK 159
 A+ +SFG LTLPHQLM+LVL+EQIYRAF I+QG PYHK
 Sbjct: 121 ANLSVSFGRITLPHQLMRLVLVEQIYRAFTIQGFPHYK 159

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5097> which encodes the amino acid sequence <SEQ ID 5098>. Analysis of this protein sequence reveals the following:

15 Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.4462(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 112/159 (70%), Positives = 133/159 (83%)

25 Query: 1 MKLKIIITVGKLKEKYLKEGVAEYQKRLNRFSKIETIELADEKTPDKASISENQRILDIEG 60
 MK+K+I VGKLKE+YLK+G++EYQKRL+RF + E IEL DE+TPDKAS ++NQ I+ E
 Sbjct: 1 MKVKLICVGKLKERYLKDGISEYQKRLSRFCQFEMIELTDERTPDKASFADNQLIMSKEA 60

30 Query: 61 ERILSKIGERDYGVLAEIGKQLPSESFSHLIDQKMISGYSTITTFVIGGSLGLSQQVKKR 120
 +RI KIGERD+VI LAIEGKQ PSE+FS LI + GYSTITF+IGGSLGL +KKR
 Sbjct: 61 QRIHKIGERDFVIALAEIGKQFPSETFSELISGVTVKGYSTITTFIIGGSLGLDSIIKKR 120

35 Query: 121 ADYLMSFGLLTLPHQLMKLVLMQIYRAFMRQGTPTYHK 159
 A+ LMSFGLLTLPHQLM+LVL EQIYRAFMI QG+PYHK
 Sbjct: 121 ANMLMSFGLLTLPHQLMRLVLTEQIYRAFMITQGSPTYHK 159

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1647

A DNA sequence (GBSx1742) was identified in *S.agalactiae* <SEQ ID 5099> which encodes the amino acid sequence <SEQ ID 5100>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3785(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1851-

Example 1648

A DNA sequence (GBSx1743) was identified in *S.agalactiae* <SEQ ID 5101> which encodes the amino acid sequence <SEQ ID 5102>. This protein is predicted to be a serine protease. Analysis of this protein sequence reveals the following:

```

5   Possible site: 29
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.4533(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9445> which encodes amino acid sequence <SEQ ID 9446> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC45334 GB:AF000658 putative serine protease [Streptococcus pneumoniae]
  Identities = 215/370 (58%), Positives = 278/370 (75%), Gaps = 20/370 (5%)

20 Query: 4  NDNIPNGGVTKTSKVNNITPTTKAVKKVQNSVSVINYKQESRSDLSDFYSHFFGNQ 63
      N++ N  +T+T+  Y N  TT+AV KV+++VVSVI Y      S      FGN
Sbjct: 46  NNSNNNSTITQTA---YKNENSTTQAVNKVKAUVSVITYSANRQNS-----VFGND 94

Query: 64  GGNTDKGLQVYEGSGVIYKKDGKNAYVVTNNHVIDGAKQIEIQLADGSKAVGKLVGSDT 123
      +TD  ++  EGSGVIYKK+ K AY+VTNNHVI+GA +++I+L+DG+K  G++VG+DT
25 Sbjct: 95  DTDTDSQ-RISSESGSGVIYKKNDKEAYIVTNNHVIINGASKVDIRLSGDTKVPGEIVGADT 153

Query: 124 YSDLAVVKIPSDKVSNAIEFADSSKLNIGETAIAIGSPLGTEYANSVTQGIVSSSLKRTVT 183
      +SD+AVVKI S+KV+ +AEF DSSKL +GETAIAIGSPLG+EYAN+VTQGIVSSL R V+
30 Sbjct: 154 FSDIAVVKISSEKVTVAEFGDSSKLTVGETAIAIGSPLGSEYANTVTQGIVSSSLNRNVS 213

Query: 184 MTNEEGQTVSTNAIQTDAAINPGNSGGALINIEGQVIGINSSKISSTSNQTSQGQSSGNSV 243
      + +E+GQ +ST AIQTD AINPGNSGG LINI+GQVIGI SSKI++      + G SV
Sbjct: 214 LKSEdGQAISTKAIQTDTAINPGNSGGPLINIQQQVIGITSSKIAT-----NGGTSV 265

35 Query: 244 EGMGFAIPSNdvVKIINQLESNGQVERPALGISMAGLSNLPsDVISKLKIPSNVTNGIVV 303
      EG+GFAIP+ND + II QLE NG+V RPALGI M  LSN+ +  I +L IPSNVT+G++V
Sbjct: 266 EGLGFAIPANDAINIIEQLEKNGKVTRPALGIQMVNLsNVSTSDIRRLNIPSNVTSGVIV 325

Query: 304 ASIQSGMPAQGKLKKYDVITKVDdKEVVSpsDLQSLLYGHQVGDsITVTFYRGENKQTVT 363
      S+QS MPA G L+KYDVITKVDdKE+ S +DLQS LY H +GD+I +T+YR  ++T +
40 Sbjct: 326 RSVQSNMPANGHLEKYDVITKVDdKEIASSTDLQsALYNHSIGDTIKITYYRNGKEETTS 385

Query: 364 IKLTkTSKDL 373
      IKL K+S DL
45 Sbjct: 386 IKLNKSSGDL 395

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5103> which encodes the amino acid sequence <SEQ ID 5104>. Analysis of this protein sequence reveals the following:

```

Possible site: 24
50 >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -8.76    Transmembrane    11 - 27 ( 6 - 31)

      ----- Final Results -----
55          bacterial membrane --- Certainty=0.4503(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 250/375 (66%), Positives = 299/375 (79%), Gaps = 5/375 (1%)

Query:	3	HNDNIPNGGVVTKTSKVNYNNTPTTKAVKKVQNSVSVSVINYKQQESRDLSDLSDFYSHFFGN	62
		H+ + N G TS + +NN T TTKAVK VQN+VVSVINY+ S S LS+ Y+ FG	
Sbjct:	34	HSPSKINSGKATTSNMVFNNNTTNTTKAVKAVQNAVSVSVINYQDNPS-SSLNFPYTKLFGE	92
Query:	63	QGG--NTDKGLQVYEGSGSVIYKKDGKNAYVVTNNHVIDGAKQIEIQLADGSKAVGKLVG	120
		N D L ++ EGSVVIY+KDG +AYVVTNNHVIDGAK+IEI +ADGSK VG+LVG	
Sbjct:	93	GRSKENKDAELSLTFSEGS SVTYRKDGNSAYVVTNNHVIDGAKRIEILMADGSKVVGELVG	152
Query:	121	SDTYSDLAVVKIPSDKVSNIAEFADSSKLNIGETAIAIGSPLGTEYANSVTQGVSSSLKR	180
		+DTYSDLAVVKI SDK+ +AEFADS+KLN+GE AIAIGSPLGT+YANSVTQGVSSSL R	
Sbjct:	153	ADTYSDLAVVKISSDKIKTVAEFADSTKLNVGGEVAIAIGSPLGTQYANSVTQGVSSSLSR	212
Query:	181	TVTMTNEEGQTVSTNAIQTDAAINPGNSGGALINIEGQVIGINSSKISSTSNQTSQGSSG	240
		TVT+ NE G+TVSTNAIQTDAAINPGNSGG LINIEGQVIGINSSKISST ++G S	
Sbjct:	213	TVTLKNENGETVSTNAIQTDAAINPGNSGGPLINIEGQVIGINSSKISSTPTGSNGNS--	270
Query:	241	NSVEGMGFAIPSNVDVVKIINQLESNGQVERPALGISMAGLSNLPSPDVISKLIKIPSNVTNG	300
		+VEG+GFAIPS DV+KII QLE+NG+V RPALGISM L++L ++ +S++ IP++VT G	
Sbjct:	271	GAVEGIGFAIPSTDVIKIIKQLETNGEVIRPALGISMVNLDLSTNALSQINIPTSVTGG	330
Query:	301	IVVASIQSGMPAQGKLKKYDVITKVDDEKVVSPSDLQSLLYGHQVGDSTITVTFYRGENKQ	360
		IVVA ++ GMPA GKL +YDVIT++D K V S SDLQS LYGH + D+I VTFYRG K+	
Sbjct:	331	IVVAEVKEGMPASGKLAQYDVITEIDGKTVNSISDLQSSLYGHDINDTIKVTFYRGTTKK	390
Query:	361	TVTIKLTKTSKDLAK	375
		IKLTKT++DL K	
Sbjct:	391	KADIKLTKTODLTK	405

```
Lipop: Possible site: -1    Crend: 10
McG: Discrim Score:      12.68
GvH: Signal Score (-7.5): -1.33
    Possible site: 21
>>> Seems to have a cleavable N-term signal seq.
ALOM program    count: 0 value: 4.56 threshold: 0.0
    PERIPHERAL Likelihood = 4.56      301
    modified ALOM score: -1.41
```

----- Final Results -----

```

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

57.4/75.6% over 386aa

Streptococcus

pneumoniae

GP|2109443| putative serine protease Insert characterized

ORF02135 (307 - 1506 of 1827)

GP|2109443|gb|AAC45334.1||AF000658(9 - 395 of 397) putative serine protease {Streptococcus pneumoniae}

%Match = 34.6

%Identity = 57.3 %Similarity = 75.6

Matches = 223 Mismatches = 89 Conservative Sub.s = 71

228 258 288 318 348 378 399 429
RLSTSCGYFLFLAFKV*LRSLSD*YKNLRR*L FVKKKLVSSLLKCSLI IIVSFAGGAFASFVMNH---NDNI PNGGVTK
 : | :: ::||| ||: || : : :: |
MEANMKHLKTFYKKWFQLLVVIVISFFSGALGSFSITQLTQKSSVMNNSNNNS

 10 20 30 40 50

-1854-

Query: 181 LTEDISVRRLKLLKQEKKTNHKSLQNKDVFLKHQENELAQFLGSKVKLTINKDGAGNIK 240
 + EDISVR+LE LL ++K+ K Q + F+++E +L + LG V++ ++K +G I
 Sbjct: 179 IEEDISVRKLEALLTEKKQ---KKQKTNHFIQNEEKQLRKLLGLDVEIKLSKKDSGKII 235

5 Query: 241 IAFANQEELNRIINTLK 257
 I+F+NQEE +RIIN+LK
 Sbjct: 236 ISFSNQEEYSRIINSLK 252

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5107> which encodes the amino acid
 10 sequence <SEQ ID 5108>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.1758(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 146/256 (57%), Positives = 191/256 (74%), Gaps = 1/256 (0%)

Query: 2 EYLETININHIAPNPYQPRLEFNTKELEELANSIKINGLIQPIIVRPSAVFGYELVAGER 61
 E L + I I NPYQPR++FN +EL++LA SIK NGLIQPIIVR S +FGYELVAGER
 Sbjct: 14 ELLIDLPIEDIVTNFYQPRIQFNQRELQDLATSIKSNGLIQPIIVRKSDFGYELVAGER 73

Query: 62 RLRAAKLAKLESIPATIKSYNNDDSMQLAIVENLQRSNLSPIEEAKAYSQLLQKKSMTHE 121
 RL+A+K+A L+ +PAIIK + +SMQ AIVENLQRSNL+ IEEAKAY L++KK MTH+
 Sbjct: 74 RLKASKMAGLKKVPATIKKISTLESMMQQAIVENLQRSNLNAIEEAKAYQLLVEKKHMTD 133

Query: 122 ELAKYMGKSRPYISNTIRLLNLPPLITSATIEEGKLSSGHARALLSLPDASQQKDWYQRI 181
 E+AKYMGKSRPYISNT+RL L LP I AIEEGK+S+GHARALL+L D QQ +I
 Sbjct: 134 ELAKYMGKSRPYISNTIRLLQLPAPIIKATIEEGKISAGHARALLTSDDKQQLYLTHKIQ 193

Query: 182 TEDISVRRLKLLKQEKKTNHKSLQNKDVFLKHQENELAQFLGSKVKLTINKDGAGNIKI 241
 E +SVR++E+L+ ++ S + K++F E +LA+ LG V + + + +G ++I
 Sbjct: 194 NEGLSVRQIEQLV-TSTPSSKLSKKTKNIFATSLEKQLAKSLGLSVNMKLTANHSGYLQI 252

Query: 242 AFANQEELNRIINTLK 257
 +F+N +ELNRIIN LK

Sbjct: 253 SFSNDELNRIINKLK 268

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1650

45 A DNA sequence (GBSx1745) was identified in *S.agalactiae* <SEQ ID 5109> which encodes the amino
 acid sequence <SEQ ID 5110>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.37 Transmembrane 2 - 18 (1 - 18)

----- Final Results -----

50 bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10297> which encodes amino acid sequence <SEQ ID
 10298> was also identified.

-1855-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5111> which encodes the amino acid sequence <SEQ ID 5112>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3646(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 353/455 (77%), Positives = 401/455 (87%), Gaps = 6/455 (1%)

Query: 32 MTENEQLFWNRVLELSRSQIAPAAYEFFVLEARLLKIEHQTAVITLDNIEMKKLFWEQNL 91
MTENEQ+FWNRVLEL++SQ+ A YEFFV +ARLLK++ A I LD +MK+LFWE+NL
Sbjct: 1 MTENEQIFWNRVLELAQSQLKQATYEFFVHDARLLKVDKHIATTIYLD--QMKELFWEKNL 58

Query: 92 GPVILTAGFEIFNAEITANYV-SNDLHLQETSFS-NYQSSNEVNTLPKIDSNLKEKY 149
VILTAGFE++NA+I+ +YV DL +++ N + +N+LP + S+L KY
Sbjct: 59 KDVILTAGFEVYNAQISVDYVFEEDLMIEQNQTAKINQPKQALNSLPT--VTSDLNSKY 116

Query: 150 TFANFVQGDENRWAVSASIAVADSPGTTYNPLFIWGGPGLGKTHLLNAIGNQVLRDNPNA 209
+F NF+QGDNRWAV+ASIAVA++PGTTYNPLFIWGGPGLGKTHLLNAIGN VL +NPNA
Sbjct: 117 SFENFIQGDENRWAVAASIAVANTPGTTYNPLFIWGGPGLGKTHLLNAIGNSVLLNPNA 176

Query: 210 RVLTYTAENFINEFVSHIRLDSMEELKEKFRNLDLLLDIDDIQSLAKKTLGGTQEEFFNTF 269
R+ YITAENFINEFV HIRLD+M+ELKEKFRNLDLLLDIDDIQSLAKKTL GTQEEFFNTF
Sbjct: 177 RIKYITAENFINEFVIHIRLDTMDLKEKFRNLDLLLDIDDIQSLAKKTLSTQEEFFNTF 236

Query: 270 NALHTNDKQIVLTSRNPQNLDLEERLVTRFSWGLFVNITPPPDFETRVAILTNKIQEYP 329
NALH N+KQIVLTSR P+ LNDLE+RLVTRF WGL VNITPPPDFETRVAILTNKIQEY
Sbjct: 237 NALHNNKQIVLTSRTPDHLNLDLRLVTRFKWGLTVNITPPPDFETRVAILTNKIQEYN 296

Query: 330 YDFPQDTIEYLAGFDSNVRELEGALKNISLVADFKHAKTITVDIAAEAIRARKNDGPIV 389
+ FPQDTIEYLAG+FDNSVR+LEGALK+ISLVA+FK TITVDIAAEAIRARK DGP +
Sbjct: 297 FIFPQDTIEYLAGQFDSNVRLLEGALKDISLVANFKQIDTITVDIAAEAIRARKQDGPKM 356

Query: 390 TVIPIEBIQVQKGYGVTVKEIKATKRTQDIVLARQVAMYLAREMTDNSLPKIGKEFGG 449
TVIPIEBIQ VQKGYGVTVKEIKATKRTQ+IVLARQVAM+LAREMTDNSLPKIGKEFGG
Sbjct: 357 TVIPIEBIQVQKGYGVTVKEIKATKRTQNIVLARQVAMFLAREMTDNSLPKIGKEFGG 416

Query: 450 RDHSTVLHAYNKIKNMVAQDDNLRIEIETIKNKIR 484
RDHSTVLHAYNKIKNM++QD++LRIEIETIKNKI+
Sbjct: 417 RDHSTVLHAYNKIKNMISQDESRLRIETIKNKIK 451

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1651

A DNA sequence (GBSx1746) was identified in *S.agalactiae* <SEQ ID 5113> which encodes the amino acid sequence <SEQ ID 5114>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0556(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

-1856-

>GP:AAC45337 GB:AF000658 beta subunit of DNA polymerase III
 [Streptococcus pneumoniae]
 Identities = 278/378 (73%), Positives = 324/378 (85%)

5 Query: 1 MIHFSINKNFFLHALTVTKRAISHKNAIPILSTVKIEVTRDAIILTGSNGQISIENTIPA 60
 MIHFSINKN FL AL +TKRAIS KNAIPILSTVKI+VT + + L GSNGQISIEN I
 Sbjct: 1 MIHFSINKNLFQALNITKRAISSKNAIPILSTVKIDVTNEGVTLIGSNGQISIENFISQ 60

10 Query: 61 SNENAGLLVTNPGSILLEAGFFINISSLPDVTLFTEIEQHQIVLTSGKSEITLKGKDV 120
 NE+AGLL+T+ GSILLEA FFIN++SSLPDVTL+F EIEQ+QIVLTSGKSEITLKGKD
 Sbjct: 61 KNEDAGLLITSLGSILLEASFFINVSSLPDVTLDKFKEIQNQIVLTSGKSEITLKGKDS 120

15 Query: 121 DQYPRLOEMTTDTPLTLETLLKLSIINETAFAASQESRPILTGVLHVISQNKYFKAVAT 180
 +QYPR+QE++ TPL LETLLK IINETAFAAS QESRPILTGVLH V+SQ+K K VAT
 Sbjct: 121 EQYPRIQEISASTPLILETKLLKKIINETAFAASTQESRPILTGVLHVISQHKELKTAT 180

20 Query: 181 DSHRMSQRTFQLEKSANNFDLVVPSKSLREFSAVFTDDIETVEVFFSDSQMLFRSENISF 240
 DSHR+SQ+ LEK++++FD+V+PS+SLREFSAVFTDDIETVE+FF+++Q+LFRSENISF
 Sbjct: 181 DSHRLSQKLTLEKNSDDFDVVPISRLREFSAVFTDDIETVEIFFANNQILFRSENISF 240

25 Query: 241 YTRLLEGNYPDTRLLTNQFETEIIFNTNALRHAMERAYLISNATQNGTVRLEIQNETVS 300
 YTRLLEGNYPDTRLL+ F T I FN LR +MERA L+S+ATQNGTV+LEI++ VS
 Sbjct: 241 YTRLLEGNYPDTRLLIPTDFNTTITFNVNLRQSMERARLLSATQNGTVKLEIKDGVVS 300

30 Query: 361 PGEDTEDFIQLITPVRTN 378
 P + EDF+QLITPVRTN
 Sbjct: 361 PADTDEDFMQILITPVRTN 378

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5115> which encodes the amino acid sequence <SEQ ID 5116>. Analysis of this protein sequence reveals the following:

35 Possible site: 14
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.70 Transmembrane 67 - 83 (67 - 83)

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 295/378 (78%), Positives = 334/378 (88%)

50 Query: 1 MIHFSINKNFFLHALTVTKRAISHKNAIPILSTVKIEVTRDAIILTGSNGQISIENTIPA 60
 MI FSIN+ F+HAL TKRAIS KNAIPILS++KIEVT + LTGSNGQISIENTIP
 Sbjct: 1 MIQFSINRTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTTLGSNGQISIENTIPV 60

55 Query: 61 SNENAGLLVTNPGSILLEAGFFINISSLPDVTLFTEIEQHQIVLTSGKSEITLKGKDV 120
 SNENAGLL+T+PG+ILLEA FFINISSLPD+++ EIEQHQ+VLTSGKSEITLKGKDV
 Sbjct: 61 SNENAGLLITSPGAILLEASFFINISSLPDISINVKEIEQHQVVLTSGKSEITLKGKDV 120

60 Query: 121 DQYPRLOEMTTDTPLTLETLLKLSIINETAFAASQESRPILTGVLHVISQNKYFKAVAT 180
 DQYPRLOE++T+ PL L+TKLLKSI IETAFAS QESRPILTGVLH+V+S +K FKAVAT
 Sbjct: 121 DQYPRLOEVSTENPLILKTKLLKSI IETAFASLQESRPILTGVLHVISNKHDFKAVAT 180

65 Query: 181 DSHRMSQRTFQLEKSANNFDLVVPSKSLREFSAVFTDDIETVEVFFSDSQMLFRSENISF 240
 DSHRMSQR L+ ++ +FD+V+PSKSLREFSAVFTDDIETVEVFFS SQ+LFRSE+ISF
 Sbjct: 181 DSHRMSQRLITLDNTSADFVVPISKSLREFSAVFTDDIETVEVFFSPSQILFRSEHISF 240

Query: 241 YTRLLEGNYPDTRLLTNQFETEIIFNTNALRHAMERAYLISNATQNGTVRLEIQNETVS 300
 YTRLLEGNYPDTRLL +FETE++FNT +LRHAMERA+LISNATQNGTV+LEI +S
 Sbjct: 241 YTRLLEGNYPDTRLLMTEFETEVEFNTQSLRHAMERAFLISNATQNGTVKLEITQNHIS 300

-1857-

Query: 301 AHVNSPEVGKVN EELDTVSLKGDLSLNISFNPTYLIESLKAVKSETVTIRFISPVRPFTLT 360
 AHVNSPEVGKVN E+LD VS G L ISFNPTYLIESLKA+KSETV I F+SPVRPFTLT
 Sbjct: 301 AHVNSPEVGKVN EELDIVSQSGDLTISFNPTYLIESLKA+KSETVKIHFLSPVRPFTLT 360

Query: 361 PGEDTEDFIQLITPVRTN 378
 PG++ E FIQLITPVRTN
 Sbjct: 361 PGDEEESFIQLITPVRTN 378

10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1652

A DNA sequence (GBSx1747) was identified in *S.agalactiae* <SEQ ID 5117> which encodes the amino acid sequence <SEQ ID 5118>. Analysis of this protein sequence reveals the following:

15 Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.0857 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10299> which encodes amino acid sequence <SEQ ID 10300> was also identified.

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC00282 GB:AF008220 YtlR [Bacillus subtilis]
 Identities = 83/298 (27%), Positives = 138/298 (45%), Gaps = 35/298 (11%)

30 Query: 19 YIIANPHAGNKNASTIVGKIQE--LVHTEDISVFYTEQKDEK--KQVINILRSFKESDH 74
 + I NP AG++N + IQ+ + + F TE + + I+ ++ +K
 Sbjct: 5 FFIINPTAGHRNGLRVWVKSQIKELIKRVEHRSFLTEHFGHAEVLARQISTIQEYKLK-R 63

35 Query: 75 LMIIGDGTLSKVMTYLPQ--HIPCTYYPVGSGNDFARALKIPNL-----KETLTA 123
 L++IGDGT+ +V+ L I ++ P G+ NDF+R I + K LT
 Sbjct: 64 LIVIGDGTMEHVVNGLKDVDIELSFVPAGAYNDFSRGFSIKKIDLIQEIKKVKRPLT- 122

40 Query: 124 IQTERLKEINCFIYDKGLIL---NSLDLGFAYVVKASNSKIKNILNRYRLGKITIYIVI 180
 +T L +N F+ DK IL N + +GF AYV KA ++ + RL + Y +
 Sbjct: 123 -RTFHLGSVN-FLQDKSQILYFMNHIGIGFDAYVNKAMEFPLRRVFLFLRLRFLVYPL- 179

45 Query: 181 AIKSLHSSK-----VQVLVEGETGQQIKLNDLYFFALANNTYFGGGITIWPKASALTA 234
 S LH+S + E ET + +D++F ++N+ ++GGG+ P A+
 Sbjct: 180 ---SHLHASATFKPFTLACTTEDETRE---PHDVWFVAVSNHPFYGGGMKAAPLANPREK 233

45 Query: 235 ELDVMVYAKGHTFLKRLSILLSLVFKRHTTSKSIKHQTFKAMTVYFPKNSLIEIDGEIV 292
 D+V + FLK+ +L + F +HT + K +T Y DGEI+
 Sbjct: 234 TFDIVIVENQPFLLKKYWLLCLMAFGKHTKMDGVTMPKAKDITFYTKDKIPFHADGEIM 291

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1653

A DNA sequence (GBSx1748) was identified in *S.agalactiae* <SEQ ID 5121> which encodes the amino acid sequence <SEQ ID 5122>. Analysis of this protein sequence reveals the following:

55 Possible site: 15
 >>> Seems to have no N-terminal signal sequence

-1858-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3792(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45338 GB:AF000658 ORFX [Streptococcus pneumoniae]
 Identities = 46/63 (73%), Positives = 57/63 (90%)

10

Query: 1 MYQVGSLEVEMKKPHACVIKETGKKANQWKVLRVGADIKIQCTNCQHVIMMSRYDFERKLK 60
 MYQVG+ VEMKKPHAC IK TGKKAN+W++ RVGADIKI+C+NC+HV+MM RYDFERK+
 Sbjct: 1 MYQVGNFVEMKKPHACTIKSTGKKANRWETRVGADIKIKCSNCEHVMMGRYDFERKMN 60

15

Query: 61 KVL 63
 K++
 Sbjct: 61 KII 63

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5123> which encodes the amino acid
 sequence <SEQ ID 5124>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25

bacterial cytoplasm --- Certainty=0.4038(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30

Identities = 63/65 (96%), Positives = 64/65 (97%)

Query: 1 MYQVGSLEVEMKKPHACVIKETGKKANQWKVLRVGADIKIQCTNCQHVIMMSRYDFERKLK 60
 MYQ+GS VEMKKPHACVIKETGKKANQWKVLRVGADIKIQCTNCQHVIMMSRYDFERKLK
 Sbjct: 1 MYQIGSFVEMKKPHACVIKETGKKANQWKVLRVGADIKIQCTNCQHVIMMSRYDFERKLK 60

35

Query: 61 KVLQP 65
 KVLQP
 Sbjct: 61 KVLQP 65

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1654

A DNA sequence (GBSx1749) was identified in *S.agalactiae* <SEQ ID 5125> which encodes the amino
 acid sequence <SEQ ID 5126>. Analysis of this protein sequence reveals the following:

45

Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.99 Transmembrane 48 - 64 (47 - 66)

----- Final Results -----

50

bacterial membrane --- Certainty=0.2996(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

55 No corresponding DNA sequence was identified in *S.pyogenes*.

-1859-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1655

A DNA sequence (GBSx1750) was identified in *S.agalactiae* <SEQ ID 5127> which encodes the amino acid sequence <SEQ ID 5128>. Analysis of this protein sequence reveals the following:

Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.4171(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1656

20 A DNA sequence (GBSx1751) was identified in *S.agalactiae* <SEQ ID 5129> which encodes the amino acid sequence <SEQ ID 5130>. This protein is predicted to be GTP-binding protein. Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.3952(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 A related GBS nucleic acid sequence <SEQ ID 8875> which encodes amino acid sequence <SEQ ID 8876> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0
McG: Discrim Score: 0.53
GvH: Signal Score (-7.5): -0.13
35 Possible site: 29
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 0 value: 1.48 threshold: 0.0
 PERIPHERAL Likelihood = 1.48 195
 modified ALOM score: -0.80

40 *** Reasoning Step: 3

 ----- Final Results -----

45 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:BAB07770 GB:AP001520 GTP-binding protein [Bacillus halodurans]
 Identities = 223/329 (67%), Positives = 273/329 (82%), Gaps = 5/329 (1%)

 Query: 1 MVEVPDERLQKLTELITPKKTVPTTFEFTDIAGIVKGASKGEGLGKFLANIREVDAIVH 60

-1860-

+VEVPD RLQKLTTEL+ PKKTVPT FEFTDIAGIV+GASKGEG LGN+FL++IR+VDAI H
 5 Sbjct: 43 IVEVPDPRLQKLTTEL VNP KKTVP TAFEFTDIAGIVEGASKGEG LGNQFLSHIRQVDAISH 102
 Query: 61 VVRAFD DENVMREQGRED AFVDPIADIDTINLELILADLESINKRYARVEKMARTQKDKE 120
 VVR FDDEN+ G VDPI DI INLELILADLES++KR++RV+K+A+T KDKE
 10 Sbjct: 103 VVRCFDDENITHVSGS---VDPIRDISVINLELILADLESVDKRFSRVQKLAKT-KDKE 157
 Query: 121 SVAEFNV LQKIKPVLEDGKSARTIEFTTEEEAKVVKGLFLLTTKPVLYVANVDEDKVADPD 180
 +VAE VL+K+K E+ K AR+IEFTTE+ K+VKGL LLT+KPVLYVANV ED V PD
 15 Sbjct: 158 AVAELEVLEK LKDAFENEKPARSIEFTTEEQQKIVKGLHLLTSKPVLYVANVSEDDVLSPD 217
 Query: 181 DIDYVNQIRAF AETENAEVVISARAE EEEISELDD EDKLEFLEAIGL TESGVDKLTAAAY 240
 D +V +++AFA EN+EV+V+SA+ EEEI+ELD E+K FLE +G+ ESG+D+L RAAAY
 20 Sbjct: 218 DNPFFVQKVKAFAAEENSEVIVVSAKIEEEIAELDGEEKAMFLEELGIQESGLDQLIRAAAY 277
 Query: 241 HLLGLGTYFTAGEKEVRAWTFKRGIKAPQAASIIHSDFERGFIRAVTMSYDDL IQYGESEK 300
 LLGL TYFTAGE+EVRAWTF++G KAPQAA IIHSDFE+GFIRA T+SY+DL++ GS
 25 Sbjct: 278 SLLGLQTYFTAGEQEVRAWTFKRGTKAPQAAGIIHSDFEKGFIRAETVSYNDLVEAGSMA 337
 Query: 301 AVKEAGRL REEGKEYIVQDGDIMEFRFNV 329
 KE G++R EGKEY+VQDGD++ FRFNV
 30 Sbjct: 338 VAKERGKVRLEGKEYVVQDGDVIHFRFNV 366

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5131> which encodes the amino acid
 25 sequence <SEQ ID 5132>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have a cleavable N-term signal seq.

30 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the databases:

>GP:BAB07770 GB:AP001520 GTP-binding protein [Bacillus halodurans]
 Identities = 259/371 (69%), Positives = 314/371 (83%), Gaps = 5/371 (1%)

40 Query: 1 MALTAGIVGLPNVGKSTLFNAITKAGAEAAANYPFATIDPNVGMVEVPDERLQKLTTELITP 60
 MALT GIVGLPNVGKSTLFNAIT+AGAE+ANYPF TIDPNVG+VEVPD RLQKLTTEL+ P
 Sbjct: 1 MALTGTGIVGLPNVGKSTLFNAITQAGAESANYPFCTIDPNVGIVEVPDRLQKLTTEL VNP 60
 Query: 61 KKTVP TTFEFTDIAGIVKGASRGEG LGNKFLANIREIDAIVHVRAFD DENVMREQGRED 120
 KKTVP T FEFTDIAGIV+GAS+GEG LGN+FL++IR++DAI HVVR FDDEN+ G
 45 Sbjct: 61 KKTVP TAFEFTDIAGIVEGASKGEG LGNQFLSHIRQVDAISHVVRFCFDDENITHVSGS-- 118
 Query: 121 AFVDPIADIDTINLELILADLESINKRYARVEKMARTQKDKE SVAEFNV LQKIKPVLEDG 180
 VDPI DI INLELILADLES++KR++RV+K+A+T KDKE+VAE VL+K+K E+
 50 Sbjct: 119 --VDPIRDISVINLELILADLESVDKRFSRVQKLAKT-KDKEAVAELEVLEK LKDAFENE 175
 Query: 181 KSARTIEFTTEDEAKVVKGLFLLTTKPVLYVANVDEDKVANPDGIDYVKQIRDF AATENAE 240
 K AR+IEFTTE++ K+VKGL LLT+KPVLYVANV ED V +PD +V++++ FAA EN+E
 55 Sbjct: 176 KPARSIEFTTEEQQKIVKGLHLLTSKPVLYVANVSEDDVLSPD DNPFFVQKVKAFAAEENSE 235
 Query: 241 VVVISARAE EEEISELDD EKEFLEAIGL TESGVDKLTAAAYHLLGLGTYFTAGEKEVRA 300
 V+V+SA+ EEEI+ELD E+K FLE +G+ ESG+D+L RAAAY LLGL TYFTAGE+EVRA
 60 Sbjct: 236 VIVVSAKIEEEIAELDGEEKAMFLEELGIQESGLDQLIRAAAYSLGLQTYFTAGEQEVRA 295
 Query: 301 WTFKRGIKAPQAAGIIHSDFERGFIRAVTMSYDDLMTYGEKAVKEAGRL REEGKEYVVQ 360
 WTF++G KAPQAAGIIHSDFE+GFIRA T+SY+DL+ GS KE G++R EGKEYVVQ
 65 Sbjct: 296 WTFKRGTKAPQAAGIIHSDFEKGFIRAETVSYNDLVEAGSMAVAKERGKVRLEGKEYVVQ 355
 Query: 361 DGDIMEFRFNV 371
 DGD++ FRFNV
 Sbjct: 356 DGDVIHFRFNV 366

-1861-

An alignment of the GAS and GBS proteins is shown below.

Identities = 316/329 (96%), Positives = 322/329 (97%)

```

5   Query: 1   MVEVPDERLQKLTELITPKKTVPTTFEFTDIAGIVKGASKGEGLGKFLANIREVDAIVH 60
      MVEVPDERLQKLTELITPKKTVPTTFEFTDIAGIVKGAS+GEGLGKFLANIRE+DAIVH
      Sbjct: 43  MVEVPDERLQKLTELITPKKTVPTTFEFTDIAGIVKGASRGEGLGKFLANIREIDAIVH 102

      Query: 61  VVRAFDDENVMREQGREDAFVDPIADIDTINLELILADLESINKRYARVEKMARTQKDKE 120
      VVRAFDDENVMREQGREDAFVDPIADIDTINLELILADLESINKRYARVEKMARTQKDKE
10   Sbjct: 103  VVRAFDDENVMREQGREDAFVDPIADIDTINLELILADLESINKRYARVEKMARTQKDKE 162

      Query: 121  SVAEFNVLQKIKPVLEDGKSARTIEFTTEEEAKVVKGLFLLTTKPVLYVANVDEDKVADPD 180
      SVAEFNVLQKIKPVLEDGKSARTIEFTE+EAKVVKGLFLLTTKPVLYVANVDEDKVA+PD
15   Sbjct: 163  SVAEFNVLQKIKPVLEDGKSARTIEFTTEDEAKVVKGLFLLTTKPVLYVANVDEDKVANPD 222

      Query: 181  DIDYVNQIRAFATENAFAEVVISARAEIEISELDDDEKLEFLEAIGLTESGVDKLTAAAY 240
      IDYV QIR FA TENAEVVISARAEIEISELDDDEK EFLEAIGLTESGVDKLTAAAY
20   Sbjct: 223  GIDYVKQIRDAATENAFAEVVISARAEIEISELDDDEKLEFLEAIGLTESGVDKLTAAAY 282

      Query: 241  HLLGLGTYFTAGEKEVRAWTFKRGIKAPQAASIIHSDFERGFIRAVTMSYDDLQYGYSEK 300
      HLLGLGTYFTAGEKEVRAWTFKRGIKAPQAA IIHSDFERGFIRAVTMSYDDL+ YGYSEK
      Sbjct: 283  HLLGLGTYFTAGEKEVRAWTFKRGIKAPQAAGIIHSDFERGFIRAVTMSYDDLMTYGYSEK 342

25   Query: 301  AVKEAGRLREEGKEYIVQDGDIMEFRFNV 329
      AVKEAGRLREEGKEY+VQDGDIMEFRFNV
      Sbjct: 343  AVKEAGRLREEGKEYVVQDGDIMEFRFNV 371

```

30 SEQ ID 8876 (GBS177) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 4; MW 41.2kDa).

The GBS177-His fusion product was purified (Figure 118A; see also Figure 202, lane 7) and used to immunise mice (lane 1 product; 20µg/mouse). The resulting antiserum was used for Western blot, FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1657

40 A DNA sequence (GBSx1752) was identified in *S.agalactiae* <SEQ ID 5133> which encodes the amino acid sequence <SEQ ID 5134>. This protein is predicted to be stage V sporulation protein C (pth). Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have no N-terminal signal sequence

```

```

45  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2212(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

50 A related GBS nucleic acid sequence <SEQ ID 10301> which encodes amino acid sequence <SEQ ID 10302> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB03787 GB:AP001507 stage V sporulation protein C
(peptidyl-tRNA hydrolase) [Bacillus halodurans]

```

-1862-

Identities = 89/187 (47%), Positives = 127/187 (67%), Gaps = 2/187 (1%)

Query: 6 VKMIVGLGNPGSKYNDTKHNIGFMAVDRIVKDLVDNFTEDKNFKAIEIGSDFINGEKIYFI 65
 +K+IVGLGNPG+KY+ T+HN+GF VD + + L++ + K G I+GEKI+ +
 5 Sbjct: 1 MKLIVGLGNPGAKYDGTNRNVGFDVVDAVARRLNIEIKQSKA-NGLYGEGRIDGEKIFLL 59

Query: 66 KPTTFMNSGIAVKALLTYYNISIKDMIYYDDLDMEVGKIRFRQKGSAGGHNGIKSIIA 125
 KP TFMN SG +V+ L YYN+ ++D+++IYDDLD+ VGKIR RQKGSAGGHNG+KS+IA
 10 Sbjct: 60 KPQTFMNRSGESVRPFLEYYNMEVEDLLVIYDDLDLPVGKIRLRQKGSAGGHNGMKSLIA 119

Query: 126 HLGTOEFDRIVKVGIGRPNRMTVINHVLGKFDKNDIEIMILNTLDKVDNAVNYLQTNDFQ 185
 HLGT +F RI+VG+ RP TV+ HVLG++ ++ I +D A + + F
 Sbjct: 120 HLGTSDFKIRRVGVDRPAPGETTVQHVLRGRYPPEEKDAISEAIDLSAEAAEAFTK-KPFL 178

Query: 186 KTMQKYN 192
 + M +N
 15 Sbjct: 179 EVMNTFN 185

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5135> which encodes the amino acid
 20 sequence <SEQ ID 5136>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.2840(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 148/189 (78%), Positives = 166/189 (87%)

Query: 5 MVKMIVGLGNPGSKYNDTKHNIGFMAVDRIVKDLVDNFTEDKNFKAIEIGSDFINGEKIYF 64
 MVKMIVGLGNPGSKY TKHNIGFMA+D IVK+LDV FT+DKNFKA+IGS FIN EK+YF
 35 Sbjct: 16 MVKMIVGLGNPGSKYEKTKHNIGFMAIDNIVKNLDVTFDDKNFKAQIGSTFINHEKVYF 75

Query: 65 IKPTTFMNSGIAVKALLTYYNISIKDMIYYDDLDMEVGKIRFRQKGSAGGHNGIKSII 124
 +KPTTFMNSGIAVKALLTYYNI I D+I+IYDDLDMEV K+R R KGSAGGHNGIKSII
 Sbjct: 76 VKPTTFMNSGIAVKALLTYYNIDITDLIVYDDLDMEVSKLRRLRSKGSAGGHNGIKSII 135

Query: 125 AHLGTOEFDRIVKVGIGRPNRMTVINHVLGKFDKNDIEIMILNTLDKVDNAVNYLQTNDF 184
 AH+GTQEF+RIKVGIGRP MTVINHV+G+F+ D I I TLD+V NAV +YLQ NDF
 40 Sbjct: 136 AHIGTQEFNRIVKVGIGRPLKGMTVINHVMGQFNTEDNIAISLTLDRVVNAVKFYLQENDF 195

Query: 185 QKTMQKYNG 193
 +KTMQK+NG
 45 Sbjct: 196 EKTMQKFNG 204

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

50 Example 1658

A DNA sequence (GBSx1753) was identified in *S.agalactiae* <SEQ ID 5137> which encodes the amino
 acid sequence <SEQ ID 5138>. This protein is predicted to be transcription-repair coupling factor (mfd).
 Analysis of this protein sequence reveals the following:

Possible site: 37
 55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2456(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1863-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAD03810 GB:AF054624 transcription-repair coupling factor
[Lactobacillus sakei]
Identities = 523/1051 (49%), Positives = 733/1051 (68%), Gaps = 20/1051 (1%)

10 Query: 1 MNIEELFSQNKVVRTWHSGLVTNSRQLVMGFSGASKAIAIASAYEKLSSKIMVVTATQTD 60
M++I + + V++ RQL+ G SG++K + +A+ Y++ + +++++ +
Sbjct: 1 MDLISMLGNTQQVQSVLENQKPGVRQLLTGLSGSAKTLFLATTYKQQRPLLIIESNMFQ 60

15 Query: 61 SDKLSSDISSLIGEDNVYQFFADDVPAEFIFSSLDKSISSRLSALRFLKPEKNGVLITS 120
+++++ D+++ + D +Y F ++V AAE SS + R+ L FL +K G+++TS
Sbjct: 61 ANQVAEDLANQLNGDQIYTFPVEEVMAAEIAVSSPESRAERVRTLSFLATGKK-GIVVTS 119

20 Query: 121 ISGLRLLLENPEVFSKSYKFEIGQECYLDKLCNVLNLGYQKVSQVFSFGESQQRGDIL 180
++G+R LLP + SQ + E+G E L L +GY + V PGEF+ RGDII+
Sbjct: 120 VAGMRRLLPTRQWRDSQTQIEMGGEVDPKILGAQLAEMGYHRDKLVGKPGEFAMRGDII 179

25 Query: 181 DIFEMTQEYPYRLEFFGDEIDGIRQFDIDTQKSLKQLESVQISPADDIILQDADFERAKK 240
DIF + E P R+E F E+D IR F+ DTQ+S++ LESV I PA D++ A E A +
Sbjct: 180 DIFPLDTENPVRIELFDTEVDAIRSFEADTQRSIENLESVAIMPATDLLANAAQLEMAE 239

30 Query: 241 KLEG-YLVTASEVQ-----RTYLSEVLSTTENHFKHSDIRRFSLIFYEKEWGI 287
L+ Y TA+++ T +S +L+ + ++ F+ Y +
Sbjct: 240 ALQADYQQTAAKITAKDDQKALAVNFETPISRLLAGE----RIENLALFVDYLYPDHTSL 295

35 Query: 288 LDYIPEGTPLFVDDFQKIVDRNAKLDEIASLLTEDLQQGKSHSSSLNYFSDPYKQLRQYQ 347
+DY + DD+ +I + L E A+ T+ L + + D + ++Q Q
Sbjct: 296 IDYFKNSGLVVFDDYPRIQETQRVLAEEAANWQTDMLGSRRLPQAQKLLVDVHMLMKQDQ 355

40 Query: 348 -PATFFSNFHKGLGNLKFDKLHHFTQYGMQEFFNQFPLLVDEINRYKKGATVLLQVDSQ 406
P + S F KG+G LK D L + +Q+FF+Q PLL E++R++K TV++ V
Sbjct: 356 HPHLYLSLFLQKGMGKLKLDLTGNMPTNRNVQFFSQMPLLKTEMSSRWQKQQQTVVVLVSDA 415

45 Query: 407 KGLNMLQENLKEYGLDLIISDKNDIVQKESQLIVGHLSNGFYFADEKIVLITEREYHRR 466
K + + + ++ ++ ++ K +V + Q++ G L NGF D K+V++TE+E+++
Sbjct: 416 KRVKKIDQTFHDFEIEATVTTTKLVAGQIQIVQGSQNGFELPDLKLVVLTEKELFNTA 475

50 Query: 467 VKRKIRRSNISNAERLKDYNELSVGDYVHVHVGKFLGIETIEIQIHRDYLTIQYQN 526
K+K+RR ++NAERLK Y+EL GDYVVH HG+G+++G+ET+E+ G+H+DY+TI Y++
Sbjct: 476 PPKKIVRRQTLANAERLKSYSSELKPGDYVHVHVGIGEYVGMETLEVDGVHGDYITILYRD 535

55 Query: 527 ADRISIPVEQIELLTKYVSADGKEPKINTLNDGRFKKAKQVAKQVEDIADDLLKLYAER 586
++ IPV Q++++ KYVSA+ K PKIN L ++K K +V+ ++EDIADDL++LYA+R
Sbjct: 536 NGKLFIPVTQLDMVQKYVSAESKTPKINKLGAEWQKTKSKVSAKIEDIADDLIELYQAR 595

60 Query: 587 SQLQGFAFSPDDNMQNDNFNDFAIVETEDQLRSIKEIKQDMEGNRPMDRLLVGDVGFQKT 646
+G+AF DD +Q DF+N FAY ET+DQLRS EIK DME RMDRLLVGDVGFQKT
Sbjct: 596 EAEKGYAFPKDDQLQADFENQFAYPETDDQLRSTAEIKHDMKVRPMDRLLVGDVGFQKT 655

65 Query: 647 EVAMRAAFKAVNDHKQVVVLVPTTVLAQQHFENFKERFSNYPVTVDVLSRFRSKKEQTD 706
EVA+RAAFKAV KQV LVPIT+LAQQH+EN RF+++PV + +LSRF+++KE T T
Sbjct: 656 EVALRAAFKAVAAGKQVAFVPTTILAQQHYENMLARFADFPVELGLLSRFKTRKEVTAT 715

Query: 707 LKRLSKGQVDIIIGTHRLLSQDVVFSDDLGLIVIDEEQRFVGVKHKELKELKTKVDVLTILT 766
LK L KGQVDI+IGTHRLLS+DVVF DLGL+++DEEQRFVGVKHKELKELK+LK+LK +VDVLTILT
Sbjct: 716 LKGLKKGQVDIVIGTHRLLSKDVVFKDLGLLIVDEEQRFVGVKHKERLKLKQKQVDVLTILT 775

Query: 767 ATPIPRTLHMSMLGIRDLSVIETPPTNRYPVQTYVLETNPGLVREAIIREIDRGQVQFVY 826
ATPIPTLHMSMLG+RDLSVIETPPTNRYP+QTYV+E N G +REAI RE++R GQVFY+
Sbjct: 776 ATPIPRTLHMSMLGVRDLSVIETPPTNRYPIQTYVMEQNAGAMREAIERELELNGQVQFYL 835

Query: 827 YNKVDTIDQKVSELQELVPEASIGFVHGQMSEIQLENTLIDFINGDYDVLVATTIETGV 886
+N+V I+Q V E+Q LVPEA++G+ HGQM+E QLE + DF+ G YDVLV TTIETGV
Sbjct: 836 HNRVSDIEQTVDEIQALVPEATVGYAHGQMTEAQLEGVYIDFVQGYDVLVTTTIETGV 895

-1864-

Query: 887 DISNVNTLFVENADHMGSLTLYQLRGRVGRSNRIAYAYLMYRPDKVLTEISEKRLDAIKG 946
 D+ NVNT+ VE+ADH GLS LYQLRGR+GRS+R+AY Y MY+PDKVLTE+SEKRL AIK
 Sbjct: 896 DMPNVNTMIVEDADHYGLSQLYQLRGRIGRSSRVAYGYFMYKPKVLTEVSEKRLQAID 955

5 Query: 947 FTELGSFGKIAMRDLISIRGAGNILGASQSGFIDSVGFEMYSQLEQAIATKQKSLIRQK 1006
 FTELGSFGKIAMRDLISIRGAGN+LG Q GFIDSVGF++YSQ+L +A+A KQK + K
 Sbjct: 956 FTELGSFGKIAMRDLISIRGAGNLLGKQHGFIIDSVGFIDLYSQMLSEAVAKKQK+VAAK 1014

10 Query: 1007 GNAELALQIDAYLPAEYISDERQKIEIYKRI 1037
 NAE+ L+++AYLP +YI+D+RQKIEIYKRI
 Sbjct: 1015 TNAEIDLKLEAYLPDDYINDQRQKIEIYKRI 1045

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5139> which encodes the amino acid sequence <SEQ ID 5140>. Analysis of this protein sequence reveals the following:

15 .Possible site: 39
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.2826(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 875/1161 (75%), Positives = 1032/1161 (88%)

25 Query: 1 MNIIELFSQNKVVRTWHSGLVTNSRQLVMGFSGASKAIAIASAYEKLSSKIMVVTATQTD 60
 M+I+ELFSQNK V++WHSGL T RQLVMG SG+SK +AIASAY KKI+VVT+TQ +
 Sbjct: 1 MDILELFSQNKVQSWHSGLTTLGRQLVMGLSGSSKTLAIASAYLDDQKKIVVVTSTQNE 60

30 Query: 61 SDKLSSDISSLIGEDNVYQFFADDVPAAEFIFSSLDKSISRLSALRFLKDPEKNGVLITS 120
 +KL+SD+SSL+ E+ V+QFFADDV AAEFIF+S+DK++SR+ L+FL++P+ GVLI S
 Sbjct: 61 VEKLASDLSSLLDEELVFQFFADDVAAAEFIFASMDKALSRIETLQFLRNPKSQGLIVS 120

35 Query: 121 ISGLRLLLPNPEVFSKSYKFEIQECYLDKLCNVLNLGYQKVSQVFSQGEFSQRGDIL 180
 +SGLR+LLPNP+VF+KSQ + +G++ D L K L+ +GYQKVSQV SPGEFS+RGDIL
 Sbjct: 121 LSGLRILLPNPDVFTKSIQLTVGEDYDSDTLTQLMTIGYQKVSQVISPGEFSRRGDIL 180

40 Query: 181 DIFEMTQEYPYRLEFFGDEIDGIRQFDIDTQKSLKQLESVQISPADDIILQDADFERAKK 240
 DI+E+TQE PYRLEFFGD+ID IRQF +TQKS +QLE + I+PA D+I + +DF+R +
 Sbjct: 181 DIYBITQELPYRLEFFGDDIDSIRQFHPETQKSFEQLEGIFINPASDLIFEVSDFQRGIE 240

45 Query: 241 KLEGYLVTADEVQRTYLSEVLSTTENHFKHSDIRRFSLIFYEKEWGILDYIPEGTPLFVD 300
 +LE L TA + +++YL +VL+ ++N FKH DIR+F S+FYEKEW +LDYIP+GTP+F D
 Sbjct: 241 QLEKALQTAQDDKKSYLELDVLAVSKNGFKHKDIRKFQSLFYKEWSSLDYIPKGTPIFFD 300

50 Query: 301 DFQKIVDRNAKLDEIASLLTLEDLQQGKSHSLNYFSDPYKQLRQYQPATFFSNFHKGLG 360
 DFQK+VD+NA+ DLEIA+LLTLEDLQQGK+ S+LNYF+D Y++LR Y+PATFFSNFHKGLG
 Sbjct: 301 DFQKLVDKNARFDLEIANLLTLEDLQQGKALSNLNYFTDNYRELRYHPATFFSNFHKGLG 360

55 Query: 361 NLKFDKLHHFTQYGMQEFFNQFPPLLVDENRYKKSGATVLLQVDSQKGLNLLQENLKEYG 420
 N+KFD++H TQY MQEFFNQFPPLL+DEI RY+K+ TV++QV+SQ L+++ ++Y
 Sbjct: 361 NIKFDQMHQLTQYAMQEFFNQFPLLIDEIKRYQKNQTTVIVQVESQYAYERLEKSFQDYQ 420

Query: 421 LDLIISDKNDIVQKESQLIVGHLSNGFYFADEKIVLITEREYHRRVKKIRRSNISNAE 480
 L + N IV +ESQ+++G +S+GFYFADEK+ LITE EIYH+++KR+ RRSNISNAE
 Sbjct: 421 FRLPLVSANQIVSRESQIVIGAISGFFYFADEKLALITEHEIYHKKIKRRARRSNISNAE 480

60 Query: 481 RLKDYNELSVGDYVVHNVHGVGKFLGIETIEIQGIHRDYLTIQYQNADRISIPVEQIELL 540
 RLKDYNEL+VG DYVVHNVHG+G+FLGIETI+IQGIHRDY+TIQYQN+DRIS+P++QI L
 Sbjct: 481 RLKDYNELAVGDYVVHNVHGVGIGRFLGIETIEIQGIHRDYVTIQYQNSDRISLPIDQISSL 540

65 Query: 541 TKYVSADGKEPKINTLNDGRFKKAKQVAKQVEDIADDLLKLYAERSQLQGFAFSPDDNM 600
 +KYVSADGKEPKIN LNDGRF+K KQ+VA+QVEDIADDLLKLYAERSQ +GF+FSPDD++
 Sbjct: 541 SKYVSADGKEPKINKLNDGRFQKTKQKVARQVEDIADDLLKLYAERSQQKGFSFSPDDDL 600

-1865-

Query: 601 QNDFDNDFAVETEDQLRSIKEIKQDMEGNRPMDRLLVGDVGFQKTEVAMRAAFKAVNDH 660
 Q FD+DFA+VETEDQLRSIKEIK DME +PMDRLLVGDVGFQKTEVAMRAAFKAVNDH
 Sbjct: 601 QRAFDDDDFAFVETEDQLRSIKEIKADMESMQPMDRLLVGDVGFQKTEVAMRAAFKAVNDH 660

5 Query: 661 KQVVVLVPTTVLAQQHFENFKERFSNYPVTVDVLSRFRSKKEQDTLTKRLSKGQVDIIIG 720
 KQV VLVPTTVLAQQH+ENFK RF NYPV VDVLSRFRSKKEQ +TL+R+ KGQ+DIIIG
 Sbjct: 661 KQVAVLVPTTVLAQQHYENFKARFENYPVEVDVLSRFRSKKEQAETLERVRKQIDIIIG 720

10 Query: 721 THRLLSQDVVFSDLGLIVIDEEQRFQGVKHKELKELKTKVDVLTLTATPIPRTLHMSMLG 780
 THRLLS+DVVFSDLGLIVIDEEQRFQGVKHKELKELKTKVDVLTLTATPIPRTLHMSMLG
 Sbjct: 721 THRLLSKDVVFSDLGLIVIDEEQRFQGVKHKELKELKTKVDVLTLTATPIPRTLHMSMLG 780

15 Query: 781 IRDLSVIETPPTNRYPVQTYVLETNPGLVREAIIREIDRGGQVFYVYNKVDITIDQKVSEL 840
 IRDLSVIETPPTNRYPVQTYVLE NPGLVREAIIRE+DRGGQ+FYVYNKVDTI++KV+EL
 Sbjct: 781 IRDLSVIETPPTNRYPVQTYVLENNPGLVREAIIREMDRGGQIFYVYNKVDITIEKKVABL 840

20 Query: 841 QELVPEASIGFVHGQMSEIQLENTLIDFINGDYDVLVATTIIEGTVDISNVNTLFVENAD 900
 QELVPEASIGFVHGQMSEIQLENTLIDFINGDYDVLVATTIIEGTVDISNVNTLF+ENAD
 Sbjct: 841 QELVPEASIGFVHGQMSEIQLENTLIDFINGDYDVLVATTIIEGTVDISNVNTLFVENAD 900

25 Query: 901 HMGLSTLYQLRGRVGRSNRIAYAYLMYRPDKVLTEISEKRLDAIKGFTELGSFGFIAMRD 960
 HMGLSTLYQLRGRVGRSNRIAYAYLMYRPDKVLTE+SEKRL+AIKGFTELGSFGFIAMRD
 Sbjct: 901 HMGLSTLYQLRGRVGRSNRIAYAYLMYRPDKVLTEVSEKRLDAIKGFTELGSFGFIAMRD 960

30 Query: 961 LSIRGAGNILGASQSGFIDSFGFEMYSQLEQAIAATKQKSLIRQKGNALALQIDAYLP 1020
 LSIRGAGNILGASQSGFIDSFGFEMYSQLEQAIA+KQK+ +RQKGN E+ LQIDAYLP
 Sbjct: 961 LSIRGAGNILGASQSGFIDSFGFEMYSQLEQAIAASKQKTTVRQKGNTEINLQIDAYLP 1020

35 Query: 1021 AEYISDERQKIETIKRIRELETRADYEALQDELIDRFGEYDPQVAYLLEIGLLKAYLDLA 1080
 +YI+DERQKI+IYKRIRE+++R DY LQDEL+DRFGEYDPQVAYLLEI LLK Y+D A
 Sbjct: 1021 DDYIADERQKIDIYKRIREIQSREDYLNQDELMDRFGEYDPQVAYLLEIALLKHYMDNA 1080

Query: 1081 FTELVERKGNEISILFEKASLKYFLTQDYFEALSQTLKARISSETNGKMEVVFNIKHKKN 1140
 F ELVERK N++ + FE SL YFLTQDYFEALSQTLK+ISE GK+++VF+++H+K+
 Sbjct: 1081 FAELVERKNNQVIVRFEVTSILTYFLTQDYFEALSQTHLKAKISEHQGKIDIVFVVRHQKD 1140

Query: 1141 YEIIIEELLKFAECFIEIKSRK 1161
 Y I+EEL+ F E EIK RK
 Sbjct: 1141 YRILEELMLFGERLSEIKIRK 1161

40

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1659

A DNA sequence (GBSx1754) was identified in *S.agalactiae* <SEQ ID 5141> which encodes the amino acid sequence <SEQ ID 5142>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.4347(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:CAB11835 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 50/84 (59%), Positives = 70/84 (82%)

Query: 1 MRLDKYLKVSRIIKRRPVAKEVADKGRVKNVGLAKSSTDLLKNDQVEIRFGNKLITVKV 60
 MRLDK+LKVSR+IKRR +AKEVAD+GR+ +NG AK+S+D+K D++ +RFG KL+TV+V
 60 Sbjct: 1 MRLDKFLKVSRLIKRRTLAKEVADQGRISINGNQAKASSDVKPGDELTVRFGQKLTVTVQV 60

Query: 61 LEMKDSTKKEDAIIKMYBIINETRI 84

-1866-

E+KD+TKKE+A MY I+ E ++
 Sbjct: 61 NELKDTTKKEEAANMYTILKEEKL 84

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5143> which encodes the amino acid sequence <SEQ ID 5144>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2963(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 72/90 (80%), Positives = 85/90 (94%)

Query: 1 MRLDKYLKVSRIIKRRPVAKEVADKGRVKVNGVLAKSSTDCLKNDQVEIRFGNKLLTVKV 60
 MRLDKYLKVSRIIKRRPVAKEVADKGRVKVNGVLAKSSTDCLKNDQVEIRFGNKLLTVKV 60
 Sbjct: 9 MRLDKYLKVSRIIKRRPVAKEVADKGRVKVNGVLAKSSTDCLKNDQVEIRFGNKLLTVKV 68

20 Query: 61 LEMKDSTKKEDAIKMYEIIINETRIETDEQA 90
 +E+KDSTKKEDA+KMYEII+ETRI +E+A
 Sbjct: 69 IEIKDSTKKEDALKMYEIISETRITLNEEA 98

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1660

A DNA sequence (GBSx1755) was identified in *S.agalactiae* <SEQ ID 5145> which encodes the amino acid sequence <SEQ ID 5146>. This protein is predicted to be DivIC homolog. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.12 Transmembrane 34 - .50 (31 - 56)

35 ----- Final Results -----

bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC98903 GB:AF023181 DivIC homolog [Listeria monocytogenes]
 Identities = 36/119 (30%), Positives = 65/119 (54%), Gaps = 2/119 (1%)

45 Query: 2 SKPNVVQLNNQYINDE-NLKKRYEAEELRRKNRLMGWVLIIFVMLLFILPTYNLVKSYRTL 60
 +K V ++ N+YI D +KK + RL +IF ++ +L T K TL
 Sbjct: 4 AKSKVARIENRYIKDTATMKKTRSRRIALFRRLAFMAIIFAVVGGLL-TITTYTKQVLT 62

50 Query: 61 QERRQEVVKLT KDYQTLTNRTENQKLLAKQLKNPDYVQKYARAKYYFSKTGEMIIPLPD 119
 +E++++ V++ K + + ++ K+L N DY+ K AR++YY SK GE+I+ +P+
 Sbjct: 63 KEKKEKQVQVDKKMVAMKDEQDSINEQIKKLHNDYIAKLARSEYYLSKDGEIIFNIPE 121

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5147> which encodes the amino acid sequence <SEQ ID 5148>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

-1867-

INTEGRAL Likelihood = -3.93 Transmembrane 34 - 50 (32 - 51)

----- Final Results -----

5 bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AAC98903 GB:AF023181 DivIC homolog [Listeria monocytogenes]
Identities = 27/116 (23%), Positives = 59/116 (50%)

Query: 3 KPSIVQLNNHYIKKENLKKKFEESQKRNRFMGWILVSMFLFILPTYNLVKSVDPEK 62
K + ++ N YIK KK R + ++ + + L T K + ++

15 Sbjct: 5 KSKVARIENRYIKDTATMKKTRSRRIALFRRLAFMAIIFAVVGGLLTITYTKQVLTLE 64

Query: 63 QNQVVKLKKKEYNELSESTKKEKQLAERLKDDNFVKKYARAKYYLSREGEIYPIPI 118
+ ++ V++ K+ + + + ++L +D+++ K AR++YYLS++GE+I+ IP

Sbjct: 65 KKEKQVQVDKKMVAMKDEQDSLNEQIKLHNDYIAKLARSEYYLSKDGEIIFNIP 120

20 An alignment of the GAS and GBS proteins is shown below.

Identities = 73/123 (59%), Positives = 99/123 (80%)

Query: 1 MSKPNVQLNNQYINDENLKKRYEAEELRRKNRLMGWVLFVMLLFILPTYNLVKSRYTL 60
M KP++VQLNN YI ENLKK++E EE +++NR MGW+L+ +M LFILPTYNLVKSRY

25 Sbjct: 1 MKKPSIVQLNNHYIKKENLKKKFEESQKRNRFMGWILVSMFLFILPTYNLVKSVDPE 60

Query: 61 QERRQEVVKLTKDYQTLNRTENQKLLAKQLKNPDYVQKYARAKYYFSKTGEMIYPLPDL 120
+++ Q+VVKL K+Y L+ T+ +K LA++LK+ ++V+KYARAKYY S+ GEMIYP+P L

30 Sbjct: 61 EKQVVKLKKKEYNELSESTKKEKQLAERLKDDNFVKKYARAKYYLSREGEIYPIPIGL 120

Query: 121 LPK 123

LPK

Sbjct: 121 LPK 123

35 SEQ ID 5146 (GBS418) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 3; MW 42kDa).

GBS418-GST was purified as shown in Figure 219, lane 4-5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1661

A DNA sequence (GBSx1756) was identified in *S.agalactiae* <SEQ ID 5149> which encodes the amino acid sequence <SEQ ID 5150>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

45

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4355(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1868-

Example 1662

A DNA sequence (GBSx1757) was identified in *S.agalactiae* <SEQ ID 5151> which encodes the amino acid sequence <SEQ ID 5152>. Analysis of this protein sequence reveals the following:

```

Possible site: 21
5  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -5.52    Transmembrane    4 - 20 ( 3 - 22)

----- Final Results -----
10      bacterial membrane --- Certainty=0.3208(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5153> which encodes the amino acid sequence <SEQ ID 5154>. Analysis of this protein sequence reveals the following:

```

Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

20 ----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

25 The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 205/428 (47%), Positives = 285/428 (65%)

30 Query: 1  MKKVLTFLLCSLYFVSIPAISTEEPLTSLQNRRYALTQTVDKEMYFDAIPERPTTKIEI 60
      M+K+L +L + + +P ISTE+ L S+N Y L Q VV +++ IP P E
      Sbjct: 1  MRKLLAAMLMTFFLTPLPVISTEKKLIFSKNAVYQLKQDVVQSTQFYNQIPSNPNLYQET 60

Query: 61  SSFQDEALTTTGETLVPNTLLSIVSLTINSNGIPVFTLSNGQFIKASREAFNDLVSKQQ 120
      +++D LT+ L N L I SL +N +PVF L++G +++A+R+ I++D+V Q
35 Sbjct: 61  CAYKSDSLTLPAGRLGVNQPLLIKSLVLNKESSLPVFELADGTYVEANRQLIYDDIVLNQV 120

Query: 121  SVSLDYWLKPSFVTYEAPYTNGVSEVKNNLKPYSRVHLVEQAETEHGIYYKTDSGFWISV 180
      + +W + Y APY G + ++ +VH + A+T HG YY D W S
40 Sbjct: 121  DIDSYFWTQKKRLYSAPYVLGTQTIPSSFLFAQKVHATQMAQTNHGTYYLIDDKGWASQ 180

Query: 181  EDLSVADNRMAKVQEVLLLEKYNKDKYGIYIKQLNTQT VAGINIDRSMYSASIAKLATLYA 240
      EDL DNRM KVQE+LL+KYN Y I++KQLNTQT AGIN D+ MY+ASI+KLA LY
50 Sbjct: 181  EDLVQFDNRMLKVQEMLLQKYNPNYSIFVKQLNTQTSAGINADKKMYAASISKLAPLYI 240

Query: 241  SQEQVKGKLSLDSKFYKDNVNQFPNSYDPSGSGKLEKKADHKLYTVKELLEATAKESD 300
      Q+Q++ KL+ + Y +VN F YDP GSGK+ K AD+K Y V++LL+A A++SD
45 Sbjct: 241  VQKQLQKKKLAENKTLTYTKDYNHFYGDYDPLGSGKISKIADNKDYRVEDLLKAVAQQSD 300

Query: 301  NVATNMLGYVNNQYDSMFQTQVDTISGMHWDMMKKRQISPQAAGKMMEAIYYQNGDIVNY 360
      NVATN+LGY+ +QYD F++++ +SG+ WDM++R ++ ++A MMEAIY+Q G I++Y
50 Sbjct: 301  NVATNILGYLCHQYDKAFRSEIKALSGIDWDMEQRLLTSRSAANMMEAIYHQKQIISY 360

Query: 361  LSKTDFDNTRIPKNIPVKVAHKIGDAYDYKHDAAIVYAEQPFIMIIFTDKSSYDDITKIA 420
      LS T+FD RI KNI V VAHKIGDAYDYKHD AIVY PFI+ IFT+KS+Y+DIT IA
55 Sbjct: 361  LSNTDFDQQRITKNITVPVAHKIGDAYDYKHDVAIVYGNTPFILSIFTNKSTYEDITAIA 420

Query: 421  DDVYQVLK 428
      DDVY +LK
60 Sbjct: 421  DDVYGILK 428

```

-1869-

SEQ ID 5152 (GBS116) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 3; MW 48.5kDa). The GBS116-His fusion product was purified (Figure 202, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 316), which confirmed that the protein is immunoaccessible on GBS bacteria.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1663

A DNA sequence (GBSx1758) was identified in *S.agalactiae* <SEQ ID 5155> which encodes the amino acid sequence <SEQ ID 5156>. Analysis of this protein sequence reveals the following:

```

10 Possible site: 28
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
15         bacterial cytoplasm --- Certainty=0.2260(Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

20 >GP: AAD35664 GB: AE001733 conserved hypothetical protein [Thermotoga maritima]
    Identities = 100/404 (24%), Positives = 181/404 (44%), Gaps = 61/404 (15%)

    Query: 19 QKVLIAVSGGIDSINLLQFLYQYQKELSSISIGIAHINHGQRKESEKEEEYIRQWQIHDV 78
              + VL+AVSGGIDS+ LL L ++ L I I AH++H R+ S ++ E++ + + ++
25 Sbjct: 6 EHVLVAVSGGIDSMILLYVLRKFSPILLKIKITAAHLDRHIRESSRRDREFVERICRQWNI 65

    Query: 79 PVFISYF-----QGIFSEDARNHRYNFFSKVMREEGYTALVTAHHADDQAETVFM 130
              G E+ AR RY+F + ++ G + + AHH +D ETV R
    Sbjct: 66 PVETSEVDVPSLWKD SGKTLEEIAREVRYDFLKR TAKKVGASKIALAHKNDLLETVVHR 125

30 Query: 131 ILRGSRLRYLSGIKQVSAFANGQLIRPFLPYKKELLP-----NIFHFEDASNASSDYLR 184
              ++RG+ L+ I + IRPFL +K+ + N+ + D +N + Y R
    Sbjct: 126 LIRGTGPLGLACISP----KREEFIRPFLVFKRSEIEEYARKNNVPYVVDETNYNV KYTR 181

35 Query: 185 NRIRNVYFPALERENNQLKDSLITLSEETECLFTALTDLTRSIEVTNCYDF----- 235
              N IR+ P ++ N ++D++ L T L + + N Y +
    Sbjct: 182 NFIRHRIVPLMKELNPTVEDAVYRLVSVTHLLRN FVERTVQDFVERN VYFYKDYAVFVEP 241

    Query: 236 --LRQTHSVQEFLLQDYISKFPDLQVSKEQFRVILKLIRTKANIDYTIKSGYFLHKDYES 293
              L V ++L++ + P+ + KLI T + + SG F+ + +
40 Sbjct: 242 EDLFLFLEVTRWVLKEMYGRVPEYE-----KLIGTLKSKRVELWSGIFVERSFGY 291

    Query: 294 FHITKIHPKTD SFKVEKRLELHNIQIFSQYLFSGYKGFISQADITIPYDT---SPIILRR 350
              + K FK + R+E+ G + I + + +R
    Sbjct: 292 VAVGK-----TVFKKKYRVEVK-----GDMLEMEGFKIRVVNNRNDMKFWVRN 334

45 Query: 351 RKEGDRIFLGNHTKKIRRLFIDEKIT--LKEREAEVIGEONKEL 392
              RKEGDRI + +K++ +FI++K+ ++R ++ E+++ L
    Sbjct: 335 RKEGDRIIVNGRERKLKDVFIKKVPTFYRDRVPLLVDDEEDRVL 378

```

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5157> which encodes the amino acid sequence <SEQ ID 5158>. Analysis of this protein sequence reveals the following:

```

    Possible site: 33
    >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2187(Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-1870-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 218/424 (51%), Positives = 290/424 (67%), Gaps = 2/424 (0%)

5 Query: 2 YNTILKDTLSKGLFTAHOQVLIIVSGGIDSLNLLQFLYQYQKELSSISIGIAHINHGQRKE 61
Y I + +K F H+ VLIIVSGG+DS+NL FLY +Q +L I IGIAH+NH QR E
Sbjct: 4 YQEIFNEIKNKAYFKNHRHVLIAVSGGVDSMNLHLFLYLFQDKLKIRIGIAHVNHKQRSE 63

10 Query: 62 SEKEEYIRQWQIHDVPVFISYFQGFSEDRARNHRYNFFSKVMREEGYTALVTAHHA 121
S+ EE Y++ W + HD+P+++S F+GIFSE AR+ RY FF +M + Y+ALVTAHH+D
Sbjct: 64 SDSEEAYLKCWAKKHDIPYVSNFEGIFSEKAARDWRVAFKSI MLKNNYSALVTAHHS 123

15 Query: 122 DQAEVFMRLIRGSRRLRYLSGIKQVSAFANGQLIRPFLPYKKELLNPFHFDASNASSD 181
DQAE+ MR++RGSRLR+LSGIK V FANGQLIRPFL + K+ LP IFHFD+SN
Sbjct: 124 DQAEITLMRLIRGSRRLRHLSGIKSVQPFANGQLIRPFLTFSKDLPEIFHFDSSNRELS 183

20 Query: 182 YLRNRIRNVYFPALERENNQLKDSLITLSEETECLFTALTDLTRSIEVINCYDFLRQTHS 241
+LRNR+RN Y P L++EN + L L+ E LF A +LT I T+ +F Q+ S
Sbjct: 184 FLNRVRNNYLP LLKQENPRFIQGLNQLALENSLLFQAFKELTNHITTTDLTEFNEQSKS 243

25 Query: 242 VQEFLLQDYISKFPDLQVSKEQFRVILKLI RTKANIDYTIKSGYFLHKDYESFHITKIHP 301
+Q FLLQDY+ FPD L + K QF +L++I+T Y +K Y++ D SF ITKI P
Sbjct: 244 IQYFLLQDYLEGFPDLDLKKSQFTQLLQIIQTAKQGYYYLKKDYIIFIDKFSFKITKIVP 303

30 Query: 302 KTDSFKVEKRLLEHNIQIFSQYLSY--GKFISQADITIPIYDTSPIILRRRKEGDRIFL 359
KT+ K EK LE + + Y FS+ Q ++IP++ S I LR R+ GD I
Sbjct: 304 KTELKVEKMLEYDSNLCYRDYFSPMPKSNEDQGVSIPLFSLSSIKLRSRQSGDYISF 363

35 Query: 360 GNHTKKIRRLFIDEKITLKEEEEAVIGEONKELIFVIVAGRTYLRKPSEHDIMKGKLYIE 419
G+ +KKIRRLFIDEK T+ ER+ A+IGEQ++++IFV++ +TYLRK +HDIM KLYI+
Sbjct: 364 GHFSKKIRRLFIDEKFTIAERQNAIIGEQDEQIIFVLIGNKTYLRKACKHDIMLAKLYID 423

Query: 420 NLEK 423
LEK
Sbjct: 424 KLEK 427

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1664

A DNA sequence (GBSx1759) was identified in *S.agalactiae* <SEQ ID 5159> which encodes the amino acid sequence <SEQ ID 5160>. This protein is predicted to be hypoxanthine-guanine phosphoribosyltransferase (hpt). Analysis of this protein sequence reveals the following:

Possible site: 50
45 >>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.32 Transmembrane 37 - 53 (37 - 53)

----- Final Results -----
50 bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:CAA48876 GB:X69123 hypoxanthine guanine
phosphoribosyltransferase [Lactococcus lactis]
Identities = 121/179 (67%), Positives = 152/179 (84%), Gaps = 1/179 (0%)

Query: 2 LENDIKKVLVSEEDIIILKTKELGAKLTADYAGKNPLLVGV LKGSVPFMAELLKHIDTHVE 61
L+ I+KVL SEE+II K+KEIG LT +Y GKNPL++G+L+GSVFP+AEL+KHID H+E
60 Sbjct: 6 LDKAIEKVLVSEEEIIEKSKELGEILTKEYEGKNPLVLGILRGSVFPFLAELIKHIDCHLE 65

-1871-

Query: 62 IDFMVSSYHGGTTSSGEVKILKDVDTNIEGRDVIFIEDIIDTGRTLKYL RDMFKYRQAN 121
 DFM VSSYHGGT SSGEVK++ DVD T ++GRD++ +EDIIDTGRTLKYL+++ ++R AN
 Sbjct: 66 TDFMTVSSYHGGTKSSGEVKLILDVD TAVKGRDILIVEDIIDTGRTLKYLKELLEHRGAN 125

Query: 122 SVKVATLFDKPEGRLVDIDADYVCYDIPNEFIVGFGLDYAENYRNL PYVGV LKEEIYSK 180
 VK+ TL DKPEG R+V+I DY + IPNEF+VGFGLDY ENYRNL PYVGV LK E+Y+K
 Sbjct: 126 -VKIVTLLDKPEG RIVEIKPDYSGFTIPNEFVVGFGLDYENYRNL PYVGV LKPEVYNK 183

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5161> which encodes the amino acid sequence <SEQ ID 5162>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4095(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 153/180 (85%), Positives = 171/180 (95%)

- Query: 1 MLENDIKKVLSEEDIIILKTKELGAKLTADYAGKNPLL VGVLKGSVPFMAELLKHIDTHV 60
 MLE DI+K+LYSE DII KTK+LG +LT DY KNPL++GVLKGSVPFMAEL+KHIDTHV
 25 Sbjct: 1 MLEQDIQKILYSENDIIRKTKKLGEQLTKDYQEK NPLMIGVLKGSVPFMAELMKHIDTHV 60
- Query: 61 EIDFMVSSYHGGTTSSGEVKILKDVDTNIEGRDVIFIEDIIDTGRTLKYL RDMFKYRQA 120
 EIDFMVSSYHGGT+SSGEVKILKDVDTNIEGRD+I +EDIIDTGRTLKYL RDMFKYR+A
 30 Sbjct: 61 EIDFMVSSYHGGTSSSGEVKILKDVDTNIEGRDIIIVEDIIDTGRTLKYL RDMFKYRKA 120
- Query: 121 NSVKVATLFDKPEGRLVDIDADYVCYDIPNEFIVGFGLDYAENYRNL PYVGV LKEEIYSK 180
 N++K+ATLFDKPEG R+V I+ADYVCY+IPNEFIVGFGLDYAENYRNL PYVGV LKEE+YSK
 Sbjct: 121 NTIKIATLFDKPEG RVVKIEADYVCYNIPNEFIVGFGLDYAENYRNL PYVGV LKEEVYSK 180

- 35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1665

- A DNA sequence (GBSx1760) was identified in *S.agalactiae* <SEQ ID 5163> which encodes the amino acid sequence <SEQ ID 5164>. This protein is predicted to be cell division protein FtsH (ftsH). Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -7.11 Transmembrane 139 - 155 (133 - 158)
 45 INTEGRAL Likelihood = -4.62 Transmembrane 8 - 24 (7 - 31)

----- Final Results -----
 bacterial membrane --- Certainty=0.3845(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC16243 GB:AF061748 cell division protein FtsH [Streptococcus pneumoniae] (ver 2)
 Identities = 490/652 (75%), Positives = 561/652 (85%), Gaps = 5/652 (0%)

- 55 Query: 5 KNNGFLKNSFIYIILLIIAVITTFQYLLKGTSSQ-NQQISYTKLVKQLKAGEIKSISYQPS 63
 +NNG +KN F+++L I ++T FQY+ G +S +QQI+YT+LV+++ G +K ++YQP+
 Sbjct: 4 QNGLIKNPFLWLLFIFFLVTFQYFYSGNNSGSGQQINYTELVQEITDGNVKELTYQPN 63

-1872-

Query: 64 GGVVEVSGTYKKAKTIKSANSFTFLGGSVATKVTGFNSVILPNDSSIKSLVSAAEENNTN 123
 G V+EVSG YK KT K F SV TKV F S ILP D+++ L A ++
 Sbjct: 64 GSVIEVSGVYKNPKTSKEGTGIQFFTPSV-TKVEKFTSTILPADTTVSELQKLATDHKAE 122

5 Query: 124 IQVKHESSSGTWISYIASFLPLVIMIGFFMMMNQGGGGGARGAMSGKNKARSSSKDEV 183
 + VKHESSSG WI+ + S +P I+ F MM GGG R MSFG++KA++++K+++
 Sbjct: 123 VTVKHESSSGIWINLLVSIVPFGILFFFLFSMMGNMGNGRNPMSFGRSKAKAANKEDI 182

10 Query: 184 KVRFSVDVAGAEKKQELIEVVDFLKDPKRYKSLGARIPAGVLLLEGPPGTGKTLAKAVAG 243
 KVRFSVDVAGAEKKQEL+EVV+FLKDPKR+ LGARIPAGVLLLEGPPGTGKTLAKAVAG
 Sbjct: 183 KVRFSVDVAGAEKKQELVEVVEFLKDPKRFTKLGARIPAGVLLLEGPPGTGKTLAKAVAG 242

15 Query: 244 EAGVPFFSISGSDFVEMFVGVGASRVRSLEFEDAKKAERAIIFIDEIDAVGRRRGAGMGGG 303
 EAGVPFFSISGSDFVEMFVGVGASRVRSLEFEDAKKA AIIIFIDEIDAVGR+RG G+GGG
 Sbjct: 243 EAGVPFFSISGSDFVEMFVGVGASRVRSLEFEDAKKAAPAIIFIDEIDAVGRQGVGLGGG 302

20 Query: 304 NDEREQTLNQLLIEMDGFEGNESIIVIAATNRSVDLDPALLRPGRFDRKVLVGPDPVKGR 363
 NDEREQTLNQLLIEMDGFEGNE IIVIAATNRSVDLDPALLRPGRFDRKVLVGPDPVKGR
 Sbjct: 303 NDEREQTLNQLLIEMDGFEGNEGIIVIAATNRSVDLDPALLRPGRFDRKVLVGPDPVKGR 362

25 Query: 364 EAILRVHAKNKPLADNVDLKVVAQQTGFGVADLENVLNEAALVAARRNKVIDASDIDE 423
 EAIL+VHAKNKPLA++VDLK+VAQQTGFGVADLENVLNEAALVAARRNK +IDASDIDE
 Sbjct: 363 EAILKVHAKNKPLAEDVDLKLVAQQTGFGVADLENVLNEAALVAARRNKSIIDASDIDE 422

30 Query: 424 AEDRVIAGPSKKDRTISERERAMVAYHEAGHTIVGLILSNARVVHKVTIVPRGRAGGYMI 483
 AEDRVIAGPSKKD+T+S++ER +VAYHEAGHTIVGL+LSNARVVHKVTIVPRGRAGGYMI
 Sbjct: 423 AEDRVIAGPSKKDKTVSQKERELVAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGGYMI 482

35 Query: 484 ALPKEDQMLLSKDDMKELAGLMGGRVAEEIIFNAQTGASNDFEQATAMARAMVTEYGM 543
 ALPKEDQMLLSK+DMKEQLAGLMGGRVAEEIIFN QTTGASNDFEQAT MARAMVTEYGM
 Sbjct: 483 ALPKEDQMLLSKEDMKELAGLMGGRVAEEIIFNVQTTGASNDFEQATQAMARAMVTEYGM 542

40 Query: 544 SEKLGVPVQYEGNHAMMAGQMSPEKSYSQAQTAQLIDDEVRLHLLNEARNKAAADIINENRDTH 603
 SEKLGVPVQYEGNHAM+ G SP+KS S QTA ID+EVR LLNEARNKAA+II NR+TH
 Sbjct: 543 SEKLGVPVQYEGNHAM-LGAQSPQKSISEQTAYEIDEVRSLLNEARNKAAEIIQSNRETH 601

Query: 604 KLIAEALLKYETLDAAQIKSIFETGKMPETENDEDKARALSYDEIKEKMQEE 655
 KLIAEALLKYETLD+ QIK+++ETGKMP E+++ ALSYDE+K KM +E
 Sbjct: 602 KLIAEALLKYETLDSTQIKALYETGKMPFAV--EEESHALSYDEVKSKMNDE 651

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5165> which encodes the amino acid sequence <SEQ ID 5166>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -7.38 Transmembrane 138 - 154 (132 - 158)

----- Final Results -----

bacterial membrane --- Certainty=0.3951(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC16243 GB:AF061748 cell division protein FtsH [Streptococcus pneumoniae] (ver 2)
 Identities = 487/654 (74%), Positives = 565/654 (85%), Gaps = 7/654 (1%)

Query: 5 KNNGFVKNSFIYILMIIVITGFGFYLGKTSTQ-SQQISYSKLIKHLKAGDIKSLSYQPS 63

+NNG +KN F+++L I ++TGFG++ G ++ SQQI+Y++L++ + G++K L+YQP+

Sbjct: 4 QNNGLIKNPFLWLLFIFFLVTFGQYFYSGNNSGSGSQINYTELVQEITDGNVKELTYQPN 63

Query: 64 GSIIEVKGKYEKPQKVTNVSGLSFLGGRASTQVTEFSSLVLPSTILKEMTAAADKNGTE 123

GS+IEV G Y+ P+ +G+ F T+V +F+S +LP+DT + E+ A + E

Sbjct: 64 GSVIEVSGVYKNPKTSKEGTGIQFFTPSV-TKVEKFTSTILPADTTVSELQKLATDHKAE 122

Query: 124 LTVKQESSSGTWITFLMSFLPIVIFAAMMMM-NQGGGGGARGAMSGKNKAKSQSKGNV 182

-1873-

+TVK ESSSG WI L+S +P I F+ MM N GGG R MSFG++KAK+ +K ++
 5
 Sbjct: 123 VTVKHESSSGIWINLLVSIVPFGILFFFLFSMMGNMGGGNRNPMSFGRSKAKAANKEDI 182
 Query: 183 KVRFTDVAGAEEEKQELVEVVDFLKNPKKYKSLGARIPAGVLLGPPGTGKTLAKAVAG 242
 KVRFDVAGAEEEKQELVEVV+FLK+PK++ LGARIPAGVLLGPPGTGKTLAKAVAG
 Sbjct: 183 KVRFSVDVAGAEEEKQELVEVVEFLKDPKRFTKLGARIPAGVLLGPPGTGKTLAKAVAG 242
 10
 Query: 243 EAGVPFFSISGSDFVEMFVGASRVRSLEFEDAKKAERAIIFIDEIDAVGRRRGAGMGGG 302
 EAGVPFFSISGSDFVEMFVGASRVRSLEFEDAKKA AIIFIDEIDAVGR+RG G+GGG
 Sbjct: 243 EAGVPFFSISGSDFVEMFVGASRVRSLEFEDAKKAAPAIIFIDEIDAVGRQGVGLGGG 302
 15
 Query: 303 NDEREQTLNQLLIEMDGFEGNENIIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDVKGR 362
 NDEREQTLNQLLIEMDGFEGNE IIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDVKGR
 Sbjct: 303 NDEREQTLNQLLIEMDGFEGNEIIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDVKGR 362
 20
 Query: 363 EAILRVHAKNKPLANDVNLKVVAQQTPGFVGADLENVLNEAALVAARRNKIKIDASDIDE 422
 EAIL+VHAKNKPLA DV+LK+VAQQTPGFVGADLENVLNEAALVAARRNK IDASDIDE
 Sbjct: 363 EAILKVHAKNKPLAEDVDLKLVAQQTPGFVGADLENVLNEAALVAARRNKSIIDASDIDE 422
 25
 Query: 423 AEDRVIAGPSKKDRITISQKEREVAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGGYMI 482
 AEDRVIAGPSKKD+T+SQKERE+VAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGGYMI
 Sbjct: 423 AEDRVIAGPSKKDKTVSQKERELVAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGGYMI 482
 30
 Query: 483 ALPKEDQMLLSKEDLKEQLAGLMGGRVAEEIVFNAQTSGASNDFEQATQIARAMVTEYGM 542
 ALPKEDQMLLSKED+KEQLAGLMGGRVAEEI+FN QT+GASNDFEQATQ+ARAMVTEYGM
 Sbjct: 483 ALPKEDQMLLSKEDMKEQLAGLMGGRVAEEIIFNVQTTGASNDFEQATQMARAMVTEYGM 542
 35
 Query: 543 SEKLGVPVQYEGNHAMPGQISPEKAYSAQTAQMIDDEVRELLNQARNQAADIINENRDTH 602
 SEKLGVPVQYEGNHAM+ Q SP+K+ S QTA ID+EVR LLN+ARN+AA+II NR+TH
 Sbjct: 543 SEKLGVPVQYEGNHAMLAGA-SPOKSISEQTAYEIDEEVRSLLNEARNKAAEIIQSNRETH 601
 Query: 603 KLIAEALLKYETLDAAQIKSIYETGKMPVDLETDDNEAHALSDEIKNKMTESE 656
 KLIAEALLKYETLD+ QIK++YETGKMP E + E+HALSYDE+K+KM + +
 Sbjct: 602 KLIAEALLKYETLDSTQIKALYETGKMP--EAVEEESHALSDEVKSKMNDEK 652

An alignment of the GAS and GBS proteins is shown below.

Identities = 550/657 (83%), Positives = 612/657 (92%), Gaps = 2/657 (0%)

40
 Query: 1 MKNNKNNGFLKNSFIYIILLIIAVITTFQYYLKGTSQSQISYTKLVKQLKAGEIKSISY 60
 MKNNKNNGF+KNSFIYIL+II VIT FQ+YLKGTS+Q+QISY+KL+K LKAG+IKS+SY
 Sbjct: 1 MKNNKNNGFVKNSFIYILMIIVITGTFQFYLKGTSQSQISYSKLIKHLKAGDIKLSY 60
 45
 Query: 61 QPSGGVVEVSGTYKAKTIKSANSFTFLGGSVATKVTGFNSVILPNDSSIKSLVSAEEN 120
 QPSG ++EV G Y+K + + + +FLGG +T+VT F+S++LP+D+ +K + +AA++N
 Sbjct: 61 QPSGSIIEVKGKYEKPQKVTVNSGLSFLGGRASTQVTEFSSSLVLPSTILKEMTAADKN 120
 50
 Query: 121 NTNIVKHESSTWISYIASFLPLVIMIGFFMMMNQGGGGARGAMSFGKNKARSSSK 180
 T + VK ESSSGTWI+++ SFLP+VI F MMMNQQGGG ARGAMSGKNKA+S SK
 Sbjct: 121 GTELTVKQESSSGTWITFLMSFLPIVIFAAMMMMNQGGGG-ARGAMSGKNKAKSQSK 179
 55
 Query: 181 DEVKVRFSVDVAGAEEEKQELIEVVDFLKDPKRYKSLGARIPAGVLLGPPGTGKTLAKA 240
 VKVRF+DVAGAEEEKQEL+EVVDFLK+PK+YKSLGARIPAGVLLGPPGTGKTLAKA
 Sbjct: 180 GNVKVRFTDVAGAEEEKQELIEVVDFLKNPKKYKSLGARIPAGVLLGPPGTGKTLAKA 239
 60
 Query: 241 VAGEAGVPFFSISGSDFVEMFVGASRVRSLEFEDAKKAERAIIFIDEIDAVGRRRGAGM 300
 VAGEAGVPFFSISGSDFVEMFVGASRVRSLEFEDAKKAERAIIFIDEIDAVGRRRGAGM
 Sbjct: 240 VAGEAGVPFFSISGSDFVEMFVGASRVRSLEFEDAKKAERAIIFIDEIDAVGRRRGAGM 299
 65
 Query: 301 GGGNDEREQTLNQLLIEMDGFEGNESIIVIAATNRSDVLDPALLRPGRFDRKVLVGPDPV 360
 GGGNDEREQTLNQLLIEMDGFEGNE+IIVIAATNRSDVLDPALLRPGRFDRKVLVGPDPV
 Sbjct: 300 GGGNDEREQTLNQLLIEMDGFEGNENIIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDV 359
 Query: 361 KGREAILRVHAKNKPLADNVDLKVVAQQTPGFVGADLENVLNEAALVAARRNKVIDASD 420
 KGREAILRVHAKNKPLA++V+LKVVAQQTPGFVGADLENVLNEAALVAARRNK IDASD
 Sbjct: 360 KGREAILRVHAKNKPLANDVNLKVVAQQTPGFVGADLENVLNEAALVAARRNKIKIDASD 419
 Query: 421 IDEAEDRVIAGPSKKDRITISERERAMVAYHEAGHTIVGLILSNARVVHKVTIVPRGRAGG 480

-1874-

```

IDEAEDRVIAGPSKKDRITIS++ER MVAYHEAGHTIVGL+LSNARVVHKVTIVPRGRAGG
Sbjct: 420 IDEAEDRVIAGPSKKDRITISQKEREMVAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGG 479

Query: 481 YMIALPKEDQMLLSKDDMKEQLAGLMGGRVAEEIIFNAQT+TGASNDFEQATAMARAMVTE 540
YMIALPKEDQMLLSK+D+KEQLAGLMGGRVAEEI+FNAQT+GASNDFEQAT +ARAMVTE
Sbjct: 480 YMIALPKEDQMLLSKEDLKEQLAGLMGGRVAEEIVFNAQTSGASNDFEQATQIARAMVTE 539

Query: 541 YGMSEKLGFPVQYEGNHAMMAGQMSPEKSYSAQTAQLIDDEVRLHLLNEARNKAADIINENR 600
YGMSEKLGFPVQYEGNHAMM GQ+SPEK+YSAQTAQ+IDDEVRL LN+ARN+AADIINENR
Sbjct: 540 YGMSEKLGFPVQYEGNHAMMPGQISPEKAYSAQTAQMIDDEVRELLNQARNQAADIINENR 599

Query: 601 DTHKLIARALLKYETLDAAQIKSIFETGKMP-ETENDEDKARALSYDEIKEKMQEED 656
DTHKLIARALLKYETLDAAQIKSI+ETGKMP + E D+++A ALSYDEIK KM E +
Sbjct: 600 DTHKLIARALLKYETLDAAQIKSIYETGKMPVDLETDDNEAHALSYDEIKNKMTESE 656

```

SEQ ID 5164 (GBS115) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 8; MW 73kDa) and in Figure 39 (lane 3; MW 73.3kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1666

A DNA sequence (GBSx1769) was identified in *S.agalactiae* <SEQ ID 5167> which encodes the amino acid sequence <SEQ ID 5168>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2983(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1667

A DNA sequence (GBSx1770) was identified in *S.agalactiae* <SEQ ID 5169> which encodes the amino acid sequence <SEQ ID 5170>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2424(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9547> which encodes amino acid sequence <SEQ ID 9548> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB12187 GB:Z99106 similar to homoserine dehydrogenase [Bacillus subtilis]
Identities = 223/448 (49%), Positives = 313/448 (69%)

```

-1875-

Query: 1 MKVVKFGGSSSLASSQQLYKVLNIIKSDYTRRFVVVSAPGKRYEEDLKMTDALIQYYQNYI 60
 MKVVKFGGSSSLAS QL KV +I+ SD R+ VVVSAPGK Y ED K+TD LI + Y+
 Sbjct: 1 MKVVKFGGSSSLASGAQLDKVFHIVTSDPARKAVVVSAPGKHYAEDTKVTDLLIACAEQYL 60

5 Query: 61 NGKDIVKDDQTWIINRYQEIIISDLGSLGSTIAEETRSIEQLASLPIENNQFLYDCFLAAGE 120
 + ++ RY I ++L LG +I E+I + L N + D A+GE
 Sbjct: 61 ATGSAPELAEAVVERYALIANELQLGQSIIEKIRDDLFTLLEGDKSNPEQYLDVAVKASGE 120

10 Query: 121 DNNAKLVATFFNQNDIPARYVHPNEAGIIVTKEPCNARIIPGSYDKIENLCLYNEVLVIP 180
 DNNAKL+A +F + A YV+P +AG+ VT EP NA+++P SY + L + +++ P
 Sbjct: 121 DNNAKLIAAYFRYKGVKA EYVNP KDAGLFVTNEPGNAQVLPESYQNLRLRERDGLIIFP 180

15 Query: 181 GFFGVTEDNQICTFSRGGSDITGSLIAAGIKADLYENFTDVGIFAHPGVVKNPHAIPE 240
 GFFG ++D + TFSR GSDITGS++A G++ADLYENFTDVD +++ +P V+NP I E
 Sbjct: 181 GFFGFSKGDVITFSRGGSDITGSLIANGLOADLYENFTDVAVSVNPSFVENPKEISE 240

20 Query: 241 LTYKEMRELAYAGFSVLHDEALLPAYRGRIPLVIKNTNPNQPGTKIVLKHTRSNIATVG 300
 LTY+EMREL+YAGFSV HDEAL+PA+R IP+ IKNTNPN GT++V K +N V G
 Sbjct: 241 LTYREMRELSYAGFSVFHDEALIPAFRAGIPVQIKNTNPNPSAEGTRVVS KRDNITNGPVG 300

25 Query: 301 IASDSRFASINVSKYLMNREVGFGRKVLQILEDLNISFEHMPTGIDDLISIVLREKELTPI 360
 IASD+ F SI +SKYLMNRE+GFGR+ LQILE+ +++EH+P+GIDD++I+LR+ ++
 Sbjct: 301 IASDTGFCISYISKYLMNREIGFGRRALQILEEHGLTYEHVPSGIDDMTII LRQGM DAA 360

30 Query: 421 GSSEVSIMFVINSKDEKRAIKALYETFF 448
 GSSEVS+MF + +E++A++ALY+ FF
 Sbjct: 421 GSSEVSMFMFGVKEAEERKAVQALYQEFF 448

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1668

A DNA sequence (GBSx1771) was identified in *S.agalactiae* <SEQ ID 5171> which encodes the amino acid sequence <SEQ ID 5172>. This protein is predicted to be CbbY family protein. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2699(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF96016 GB:AE004353 CbbY family protein [Vibrio cholerae]
 Identities = 59/190 (31%), Positives = 93/190 (48%), Gaps = 10/190 (5%)

50 Query: 4 YKAIIFDMDGVLFDTELFYKRRERFLKQHGITIDHLP MNFFIGGNMKQVWKSVLGDQYD 63
 ++A IFDMDG+L DTE + + G+ IG N K + +L Y
 Sbjct: 6 FQAIFDMDGLLLDTERVCMRVFQEA CTACGLPFRQEVYLSVIGCNAKTI-NGILSQAYG 64

55 Query: 64 TWDIDKL----QQDYSRYKEDNPLPYKDLIFQDCKRVIEKLHHKGYLLGLASSSTRHDIM 119
 D+ +L +Q Y+ +P+KD + ++E L + + +A+S+ + +
 Sbjct: 65 E-DLPR LHNEWQRYNNAVVMHEAIPHKG DVIA----LLEWLKARSIPVAVATSTQKEVAL 119

60 Query: 120 LALESFNLDTYFKVILSGEEFSES KPNPAIYNRAAELLDPKQILIVEDSEKGITAGIA 179

-1876-

+ L+ LD YF I +G E ++ KP+P IY AAE L + QQ L EDS GI A +A
 Sbjct: 120 IKLQLAGLDHYFANITTGCEVTQGKPHPEIYLLAAERLGVPEQQCLAFEDSNNGIKAAMA 179
 Query: 180 AGIDVWAIED 189
 A + + I D
 Sbjct: 180 AQMHAFQIPD 189

There is also homology to SEQ ID 448.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1669

A DNA sequence (GBSx1772) was identified in *S.agalactiae* <SEQ ID 5173> which encodes the amino acid sequence <SEQ ID 5174>. This protein is predicted to be Pseudomonas putida enoyl-CoA hydratase II homologue (b1394). Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -2.18 Transmembrane 128 - 144 (128 - 145)
 INTEGRAL Likelihood = -1.06 Transmembrane 154 - 170 (154 - 170)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9549> which encodes amino acid sequence <SEQ ID 9550> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5175> which encodes the amino acid sequence <SEQ ID 5176>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.08 Transmembrane 110 - 126 (109 - 128)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2232(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 150/263 (57%), Positives = 197/263 (74%)
 Query: 19 LKFENIIYGIDGNVATIMLNRPDISNGFNIPMCQEIIDAIRLVSENKDVMLVIEAQGP 78
 ++F++II+ + ++AT+ LNRP++SNGFNIP+CQEI+ A+ V + V FL+I+A G +
 Sbjct: 1 MQFKHIIIFDVDDLATLTLNRPEVSNGFNIPICQEILVALAEVKRDTSVRFLLIKAVGKV 60
 Query: 79 FSIGGDLKVMKAAVESDDISSLTKEIAELVNQISYDLLQLEKPVVMCVDGAVAGAAANIAL 138
 FS+GGDL M+ AV D++ SL KLAELV +IS+ + L KPV++C DGAVAGAA NIAL
 Sbjct: 61 FSVGGDLVEMQEAVAKDNVQSLVKIAELVQEISFAIKHLPKPVILCADGAVAGAAFNIAL 120
 Query: 139 AADFVIASKSKKFIQAFVGVGLAPDAGGLLLLSKSGITRAVQLALTGESLSAEKAEALG 198
 A DF IAS ++KFIQAFV VGLAPDAGGL LL++++G+ RA L +TGE ++A+K G
 Sbjct: 121 AVDFCIASQTQKFIQAFVNVGLAPDAGGLFLLTRAVGLNRATHLVMTGEGITADKGLDYG 180
 Query: 199 IVYKLCESDKIGKIKDQLLKRLSRHSINSYQAISLAWEEAFKDWQYKKLELQIQESLA 258
 VY+ ESDK+ K+ QLLKRL R S NSY +KSL W++ F WE Y K EL +QE LA
 Sbjct: 181 FVYRTAESDKLDKVCQLLKRLRRGSSNSYAGMKSLVWQSFFTGWEDYAKAELAIQEELA 240

Query: 259 FKQDFKEGVRAHADRRRPNFLGK 281
FK+DFKEGV A +RRRPNF GK
Sbjct: 241 FKEDFKEGVIAFGERRRPNFOGK 263

```
Lipop: Possible site: -1    Crend: 10
SRCFLG: 0
McG: Length of UR: 9
      Peak Value of UR: 1.45
      Net Charge of CR: -1
McG: Discrim Score: -5.99
GvH: Signal Score (-7.5): -4.37
      Possible site: 27
```

Amino Acid Composition: calculated from 1

INTEGRAL Likelihood = -2.18 Transmembrane 110 - 126 (110 - 127)

INTEGRAL Likelihood = -1.06 Transmembrane 136 - 152 (136 - 152)

PERIPHERAL Likelihood = 1.32' 49

modified ALOM score: 0.94

icm1 HYPID: 7 CFP: 0.187

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.1871(Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

ORF01047 (355 - 1143 of 1443)

GP|3253198|gb|AAC24330.1||AF029714 (1 - 263 of 263) PhaB {Pseudomonas putida}

```
%Match = 15.4
```

%Identity = 33.3 %Similarity = 56.4

Matches = 88 Mismatches = 113 Conservative Sub.s = 61

96 126 156 186 216 246 276 306
*KTVRRGLQVLVOPVLMCGLLKINTLE*ISRRLMY**AI*VNFL*N*ITIKNGKENSVFLEFILP*KLGL**NTKHDNLI

336 366 396 426 456 486 516 546
TKLFFELSLIKRGDILKFENIIYGIDGNVATIMLNRPDISNGFNIPMCORIIDAIRLVSENKDVMLFLVIEAOGPIFSIG

```

: |::|:: | : || : |||: | || | | : ::: |:: | |:: | | |
MTFOHILFSIEDGVAFLSLNRPEOLNSEFNAAMHLEVREALKOVROSSDARVLLLTAEGRGFCAG

```

10 20 30 40 50 60

576 606 636 666 696 726 756 786
GDI.KVMKAAVESDDISSLT.KTIAELVNOISYDPLLEKPVVMCVDGAVAGAAANTALAADPVIASKKSKFIOAFVGVGLAP

ONT.SDRNVA PDAEVYDPI.GESINKEVNDI.VPPT.PDI.PI.PVT.CAVNGVAAGAGANI.PI.ACI.VI.AGRSASETOAECKTGLVYD

80 90 100 110 120 130 140

816 846 876 906 936 966 996 1026

DACCIIILSKSGITRAVOLATGESISAEKARALGIVYKICSDKIGKTKDOLIKRISPHSINGVOATKSLAWTAARKD

DOCCITWIL.DPT.YQMADAPALAMICEPTICAEPOACOWCTIUDIVONDAAL.PDEALT.TT.ABOI.ACQDTYVCTALIK_PSLTASAFNM

DSGGTWEEFRKVGMAKAKAHLHMLGKNGAEQAQQWGLTHRVVDDAALNDEALIHARQLASQFIIGDALIRKSLNAS
160 170 180 180 200 210 220

1053 1083 1113 1143 1173 1203 1233 1263

MEQYKLELOQESTAFKODKFCVBAHADPPDNFIS*FENOTT*DSLANKEFI*VNLITK*GVVTSWATPIET

-WEQIRKEELQZQESLAR KQDFKEGVRAHADRKRNF EGR
 :: :|| || :||::||| | ::| | | :

RLAGRSEDYREGVSAFMNKRTPAE

240 250 260

-1878-

SEQ ID 8878 (GBS374) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 8; MW 32kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 71 (lane 2; MW 57kDa).

The GBS374-GST fusion product was purified (Figure 215, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 307), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1670

A DNA sequence (GBSx1773) was identified in *S.agalactiae* <SEQ ID 5177> which encodes the amino acid sequence <SEQ ID 5178>. This protein is predicted to be a 16.1 kDa transcriptional regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD05186 GB:AF110185 unknown [Burkholderia pseudomallei]
Identities = 30/102 (29%), Positives = 60/102 (58%)

Query: 32 DVSLKEMHTIEIIGKHSEVTPSDVARELMLTLGTVTSLNKLEKKGYIERKRSSIDRRVV 91
      +++ +++ I ++ + TP +++R+L G++T L++LEKKG++ R RS DRRV+
Sbjct: 39 ELTAQQISVILLIARGYARTPFELSRKLSYDSGSMTRMLDRLEKKGFFVVRARSESIDRRVI 98

Query: 92 HLSLTKRGRLLDRLHSEFKHSMVSHIIEDLGEEDIKMLTSAL 133
      L+LT+RG R + ++ +E +++ +LT L
Sbjct: 99 ELALTERGAHAARALPALIATELNAQLEGFSADELALLTDL 140
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5179> which encodes the amino acid sequence <SEQ ID 5180>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1412(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 111/144 (77%), Positives = 129/144 (89%)

Query: 1 MEYDQINSYLVDIFNRIIMIEEMSLKTSQFSDVSLKEMHTIEIIGKHSEVTPSDVARELM 60
      +EYD+I YLVDIFNRI++IEEMSLKTSQFSDVSLKEMHTIEIIGK+ +VTPSD+ARELM
Sbjct: 7 LEYDKIYPYLVDIFNRIILVIEEMSLKTSQFSDVSLKEMHTIEIIGKYDQVTPSDIARELM 66

Query: 61 LTLGTVTTSNLKLEKKGYIERKRSSIDRRVVHLSLTKRGRLLDRLHSEFKHSMVSHIIED 120
      +TLGTVTTSNLKLE KGYI R RS DRRVV+LSLTKRGRLLDRLH+KFHK+MV H+I D
Sbjct: 67 VTLGTVTTSNLKLEAKGYIARTRSRSDRRVVYLSLTKRGRLLDRLHAKFHKNMVGHVIAD 126
```


-1879-

Query: 121 LGSEEDIKMLTSALGNLHKFLEDLV 144
 + +E+++ L LGNLH+FLEDLV
 Sbjct: 127 MSDEEMQALVRGLGNLHQFLEDLV 150

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1671

- A DNA sequence (GBSx1774) was identified in *S.agalactiae* <SEQ ID 5181> which encodes the amino acid sequence <SEQ ID 5182>. This protein is predicted to be 3-oxoacyl-(acyl-carrier-protein) synthase III (fabH-2). Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.12 Transmembrane 103 - 119 (103 - 119)

- 15 ----- Final Results -----
 bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF98271 GB:AF197933 beta-ketoacyl-ACP synthase III
 [Streptococcus pneumoniae]
 Identities = 225/324 (69%), Positives = 276/324 (84%), Gaps = 1/324 (0%)

- 25 Query: 1 MVFAKISQLAHYAPSQIIKNEEDLSLIMDTSDDWISSRTGIKQRHISKNETTADLANKVAE 60
 M FAKISQ+AHY P Q++ N DL+ IMDT+D+WISSRTGI+QRHIS+ E+T+DLA +VA+
 Sbjct: 1 MAFAKISQVAHYVPEQVVNHDLAQIMDTNDEWISSRTGIRQRHISRTESTSDLATEVAK 60
- 30 Query: 61 QLIEKSGYSASQIDFIIIVATMTDPSMMPSTAARVQAHIGASNAFAFDLSAACSGFVFALS 120
 +L+ K+G + ++DFII+AT+TPDSMMPSTAARVQA+IGA+ AFAFDL+AACSGFVFALS
 Sbjct: 61 KLMAKAGITGEELDFIILATITTPDSMMPSTAARVQANIGANKAFAFDLTAACSGFVFALS 120
- 35 Query: 121 TAEKLISGSYQKGLVIGAETVSKVLDWTDRTAVLFGDGAGGVLEASKEKHFLAESLN 180
 TAEK I+SG +QKGLVIG+ET+SK +DW+DR TAVLFGDGAGGVLEAS+++HFLAESLN
 Sbjct: 121 TAEKFIASGRFQKGLVIGSETLSKAVDWSRSTAVLFGDGAGGVLEASEQEHFLAESLN 180
- 40 Query: 181 TDGSR-QGLQSSQVGLNSPFSDEVLDKFLKMDGRAIFDFAIKEVSKSINHLIETSYLEK 239
 +DGSR + L GL+SPFSD+ D FLKMDGR +FDFAI++V+KSI I+ S +E
 Sbjct: 181 SDGSRSECLTYGHSGHLSPFSQDESADSFLLKMDGRTVDFFAIRDVAKSIQTIDESPIEV 240
- 45 Query: 240 EDIDYLFHLQANRRILDKMSRKIDDIARDKFPENMMDYGNTSAASIPILLSESYENGLLKL 299
 D+DYL LHQAN RILDKM+RKI + R K P NMM+YGNTSAASIPILLSE E GL+ L
 Sbjct: 241 TDLDYLLHLQANDRIIDKMARKIGVDRAKLPANMMEYGNTSAASIPILLSECVEQGLIPL 300
- Query: 300 DGNQTILLSGFGGGLTWGSLIVKI 323
 DG+QT+LLSGFGGGLTWG+LI+ I
 Sbjct: 301 DGSQTVLLSGFGGGLTWGTLILTI 324

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5183> which encodes the amino acid sequence <SEQ ID 5184>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.16 Transmembrane 103 - 119 (103 - 120)

- 55 ----- Final Results -----
 bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1880-

The protein has homology with the following sequences in the databases:

>GP:AAF98271 GB:AF197933 beta-ketoacyl-ACP synthase III

[Streptococcus pneumoniae]

Identities = 212/324 (65%), Positives = 263/324 (80%)

5 Query: 1 MIFSKISQVAHYVPQQLVTNNDLASIMDTSEWIFSRGTGIAERHISRDEMTSDLAIQVAD 60
M F+KISQVAHYVP+Q+VTN+DLA IMDT+ EWI SRTGI +RHIS E TSDLA +VA
Sbjct: 1 MAFKISQVAHYVPEQVVINHDLAQIMDTNDEWISSRTGIRQRHISRTESTSDLAATEVAK 60

10 Query: 61 QLLTQSGLKADAIDFIIVATISPDATMPSTAARKVQAAIAATSFAFADMTAACSGFVFALA 120
+L+ ++G+ + +DFII+ATI+PD+ MPSTAA+VQA I A AFAFD+TAACSGFVFAL+
Sbjct: 61 KLMAKAGITGEELDFIILATITPDSPMPSTAARVQANIGANKAFADFALTAACSGFVFALS 120

15 Query: 121 MADKLIASGAYQNGMVIGAETLSKLVNWQDRATAVLFGDGAGGVILLEASKDKHVLAEHL 180
A+K IASG +Q G+VIG+ETLSK V+W DR+TAVLFGDGAGGVILLEAS+ +H LAE+L+
Sbjct: 121 TAEKFIASGRFQKGLVIGSETLSKAVDWSRSTAVLFGDGAGGVILLEASEQEHFLAESLN 180

20 Query: 181 TDGARCQSLISGETSLSSPYSIGKKAIAITIQMDGRAIFDFAIRDVSKSILTLMAQSDITK 240
+DG+R + L G + L SP+S + A + ++MDGR +FDFAIRDV+KSI + +S I
Sbjct: 181 SDGSRSECLTYGHSGLHSPFSDQESADSLKMDGRTVDFFAIRDVAKSIQTIDESPIEV 240

25 Query: 241 DDIDYCLLHQANRRILDKIARKIDVPREKFLNMMRYGNTSAASIPILLSEAVQKQIRL 300
D+DY LLHQAN RILDK+ARKI V R K NMM YGNTSAASIPILLSE V++G I L
Sbjct: 241 TDLDYLLHQANDRILDKMARKIGVDRAKLPANMMEYGNTSAASIPILLSECVQGLIPL 300

30 Query: 301 DGTQKILLSGFGGGLTWGSLIVRI 324
DG+Q +LLSGFGGGLTWG+LI+ I
Sbjct: 301 DGSQTVLLSGFGGGLTWGTLILTI 324

An alignment of the GAS and GBS proteins is shown below.

Identities = 216/324 (66%), Positives = 271/324 (82%), Gaps = 1/324 (0%)

35 Query: 1 MVFAKISQLAHYAPSQIIKNEEDLSLIMDTSDDWISSRTGIKQRHISKNETTADLANKVAE 60
M+F+KISQ+AHY P Q++ N DL+ IMDT+ +WI SRTGI +RHIS++E T+DLA +VA+
Sbjct: 1 MIFSKISQVAHYVPQQLVTNNDLASIMDTSEWIFSRGTGIAERHISRDEMTSDLAIQVAD 60

40 Query: 61 QLIEKSGYSASQIDFIIVATMTPDSPMPSTAARVQAHIGASNAFAFDLSAACSGFVFALS 120
QL+ +SG A IDFIIVAT++PD+ MPSTAA+VQA I A++AFAFD++AACSGFVFAL+
Sbjct: 61 QLLTQSGLKADAIDFIIVATISPDATMPSTAARKVQAAIAATSFAFADMTAACSGFVFALA 120

45 Query: 121 TAEKLISGSYQKGLVIGAETVSKVLDWTDRTAVLFGDGAGGVILLEASKEKHFLAESLN 180
A+KLI+SG+YQ G+VIGAET+SK+++W DR TAVLFGDGAGGVILLEASK+KH LAE+L+
Sbjct: 121 MADKLIASGAYQNGMVIGAETLSKLVNWQDRATAVLFGDGAGGVILLEASKDKHVLAEHL 180

50 Query: 181 TDGSR-QGLQSSQVGLNSPFSDEVLDDKFLKMDGRAIFDFAIKEVSKSINHLETSYLEK 239
TDG+R Q L S + L+SP+S ++MDGRAIFDFAI++VSKSI L+ S + K
Sbjct: 181 TDGARCQSLISGETSLSSPYSIGKKAIAITIQMDGRAIFDFAIRDVSKSILTLMAQSDITK 240

55 Query: 240 EDIDYFLHQANRRILDKMSRKIDIAKDFPENMDYGNTSAASIPILLSESYENGLLKL 299
+DIDY LHQANRRILDK++RKID+ R+KF ENMM YGNTSAASIPILLSE+ + G ++L
Sbjct: 241 DDIDYCLLHQANRRILDKIARKIDVPREKFLNMMRYGNTSAASIPILLSEAVQKQIRL 300

Query: 300 DGNQTILLSGFGGGLTWGSLIVKI 323
DG Q ILLSGFGGGLTWGSLIV+I
Sbjct: 301 DGTQKILLSGFGGGLTWGSLIVRI 324

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1881-

Example 1672

A DNA sequence (GBSx1775) was identified in *S.agalactiae* <SEQ ID 5185> which encodes the amino acid sequence <SEQ ID 5186>. This protein is predicted to be acyl carrier protein (acpP). Analysis of this protein sequence reveals the following:

```

5   Possible site: 59
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3083(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9551> which encodes amino acid sequence <SEQ ID 9552> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:AAF98272 GB:AF197933 acyl carrier protein [Streptococcus pneumoniae]
    Identities = 64/74 (86%), Positives = 67/74 (90%)

    Query: 17 MAVFEKVQEIIVEELGKDAEEVTLNNTTFDDLDADSLDVFQVISEIEDAFDIQIETEEGLN 76
20      MAVFEKVQEIIVEELGKDA EVTL +TFDDLDADSLD+FQVISEIEDAFDIQIE E L
    Sbjct: 1 MAVFEKVQEIIVEELGKDASEVTLESTTFDDLDADSLDLFQVISEIEDAFDIQIEAENDLK 60

    Query: 77 TVGDLVAYVEEKVK 90
      TVGDLVAYVEE+ K
25    Sbjct: 61 TVGDLVAYVEEQAK 74

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5187> which encodes the amino acid sequence <SEQ ID 5188>. Analysis of this protein sequence reveals the following:

```

30   Possible site: 43
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2995(Affirmative) < succ>
35      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 70/74 (94%), Positives = 71/74 (95%)

40   Query: 17 MAVFEKVQEIIVEELGKDAEEVTLNNTTFDDLDADSLDVFQVISEIEDAFDIQIETEEGLN 76
      MAVFEKVQEIIVEELGK+ EEVTL TTFDDLDADSLDVFQVISEIEDAFDIQIETEEGLN
    Sbjct: 1 MAVFEKVQEIIVEELGKETEEVTLNNTTFDDLDADSLDVFQVISEIEDAFDIQIETEEGLN 60

    Query: 77 TVGDLVAYVEEKVK 90
45      TVGDLVAYVEEK K
    Sbjct: 61 TVGDLVAYVEEKSK 74

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1673

A DNA sequence (GBSx1777) was identified in *S.agalactiae* <SEQ ID 5189> which encodes the amino acid sequence <SEQ ID 5190>. Analysis of this protein sequence reveals the following:

```

    Possible site: 31
    >>> Seems to have no N-terminal signal sequence

```

-1882-

INTEGRAL Likelihood = -0.27 Transmembrane 156 - 172 (156 - 173)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAF98273 GB:AF197933 trans-2-enoyl-ACP reductase II
[Streptococcus pneumoniae]
Identities = 257/318 (80%), Positives = 277/318 (86%), Gaps = 1/318 (0%)

Query: 1 MKTRITELLNKYPFIQGGMAWVADGDLGAVSKAGGLGIIGGNAPKEVVKANIDKIKS 60
15 Sbjct: 1 MKTRITELLNKIDYPFIQGGMAWVADGDLGAVSKAGGLGIIGGNAPKEVVKANIDKIKS 60

Query: 61 MTDKPFQGVNIMLLSPFVDDIVDLVIEEGVKVVTGAGNPGKYMERFHEAGITVIPVPSV 120
+TDKPFQGVNIMLLSPFV+DIVDLVIEEGVKVVTGAGNPGKYMERFHEAGI VIPVPSV
20 Sbjct: 61 LTDKPFQGVNIMLLSPFVEDIVDLVIEEGVKVVTGAGNPGKYMERFHEAGIIVIPVPSV 120

Query: 121 ALAKRMEKLGADAIITEGMEAGGHIGKLTMTLVRQVDAVTIPVIAAGGIADGRGAAAG 180
ALAKRMEK+GADA+I EGMEAGGHIGKLTMTLVRQV A++IPVIAAGGIADG GAAAG
25 Sbjct: 121 ALAKRMEKIGADAVIAEGMEAGGHIGKLTMTLVRQVATAISIPVIAAGGIADGEGAAAG 180

Query: 181 FMLGADAVQVGTRFVVAKESSNAHPNYKAKILKAKDIDTAVSAQVVGHPVRALKNKLVTY 240
FMLGA+AVQVGTRFVVAKESSNAHPNYK KILKA+DIDT +SAQ GH VRA+KN+L +
30 Sbjct: 181 FMLGAEAVQVGTRFVVAKESSNAHPNYKEKILKARDIDTTISAQHFGHAVRAIKNQLTRDF 240

Query: 241 SQAEDYLAGRISINEI-EELGAGALRNAVVDGVDVINGSVMAGQIAGLIKSEETCQEILE 299
AEKD EI E++GAGAL AVV GDV GSVMAQIAGL+ EET +EIL+
35 Sbjct: 241 ELAEKDAFKQEDPDLEIFEQMGAGALAKAVVHGDVDGGSVMAQIAGLVSKEETAEEILK 300

Query: 300 DIYSGARQVILSEASRWS 317
D+Y GA +.I EASRW+
Sbjct: 301 DLYYGAACKIQEEASRWT 318

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5191> which encodes the amino acid sequence <SEQ ID 5192>. Analysis of this protein sequence reveals the following:

40 Possible site: 35
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.70 Transmembrane 106 - 122 (106 - 124)
INTEGRAL Likelihood = -0.22 Transmembrane 156 - 172 (156 - 173)

45 ----- Final Results -----
bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

50 >GP:AAF98273 GB:AF197933 trans-2-enoyl-ACP reductase II
[Streptococcus pneumoniae]
Identities = 252/320 (78%), Positives = 276/320 (85%), Gaps = 1/320 (0%)

55 Query: 1 MKTRITELLNIDYPFIQGGMAWVADGDLGAVSNAGGLGIIGGNAPKEVVKANIDRVKA 60
MKTRITELLNIDYPFIQGGMAWVADGDLGAVS AGGLGIIGGNAPKEVVKANID++K+
Sbjct: 1 MKTRITELLNKIDYPFIQGGMAWVADGDLGAVSKAGGLGIIGGNAPKEVVKANIDKIKS 60

Query: 61 ITDRPFQGVNIMLLSPFADDDIVDLVIEEGVKVVTGAGNPGKYMERLHQAGIIVPVVPSV 120
+TD+PFGVNIMLLSPF +DIVDLVIEEGVKVVTGAGNPGKYMER H+AGIIV+PVVPSV
60 Sbjct: 61 LTDKPFQGVNIMLLSPFVEDIVDLVIEEGVKVVTGAGNPGKYMERFHEAGIIVIPVPSV 120

Query: 121 ALAKRMEKLGVDVAVIAEGMEAGGHIGKLTMTSLVRQVVEAVSIPVIAAGGIADGHGAAAA 180
ALAKRMEK+G DAVIAEGMEAGGHIGKLTMT+LVRQV A+SIPVIAAGGIADG GAAA
Sbjct: 121 ALAKRMEKIGADAVIAEGMEAGGHIGKLTMTLVRQVATAISIPVIAAGGIADGEGAAAG 180

-1883-

Query: 181 FMLGAEAVQIGTRFVAKESNAHQNFKDILAAKDIDTIVISAQVVGHPVRSIKNKLT SAY 240
 FMLGAEAVQ+GTRFVAKESNAH N+K+KIL A+DIDT ISAQ GH VR+IKN+LT +
 Sbjct: 181 FMLGAEAVQVGRFVAKESNAHPNYKEKILKARDIDTTISAQHFGHAVRAIKNQLTRDF 240

Query: 241 AKA EK-AFLIGQKTATDIEEMGAGSLRHAVIEGDVVGNSVMAGQIAGLVRKEESCETILK 299
 AEK AF E+MGAG+L AV+ GDV GSV MAGQIAGLV KEE+ E ILK
 Sbjct: 241 ELAEKDAFKQEDPDLEIFEQMGAGALAKAVVHGVDVGGSVMAGQIAGLVSKETAEBILK 300

Query: 300 DIYYGAARVIQNEAKRWQSV 319
 D+YYGAA+ IQ EA RW V
 Sbjct: 301 DLYYGAAKKIQEEASRWTV 320

An alignment of the GAS and GBS proteins is shown below.

Identities = 253/319 (79%), Positives = 291/319 (90%)

Query: 1 MKTRITELLNIKYPFQGGMAWVADGDLGAVSKAGGLGIIGGNAPKEVVKANIDKIKS 60
 MKTRITELLNI YPIFQGGMAWVADGDLGAVS AGGLGIIGGNAPKEVVKANID++K+
 Sbjct: 1 MKTRITELLNIDYPIFQGGMAWVADGDLGAVSNAGGLGIIGGNAPKEVVKANIDRVKA 60

Query: 61 MTDKPFQGVNIMLLSPFVDDIVDLVIEEGVKVVTGAGNPGKYMERFHEAGITVIPVPSV 120
 +TD+PFGVNIMLLSPF DDIVDLVIEEGVKVVTGAGNPGKYMER H+AGI V+PVVPSV
 Sbjct: 61 ITDRPFQGVNIMLLSPFADDIVDLVIEEGVKVVTGAGNPGKYMERLHQAGIIVPVVPSV 120

Query: 121 ALAKRMEKLGADAIITEGMEAGGHIGKLTMTLVRQVVDVAVTIPVIAAGGIADGRGAAAG 180
 ALAKRMEKLG DA+I EGMEAGGHIGKLTMT+LVRQVV+AV+IPVIAAGGIADG GAAA
 Sbjct: 121 ALAKRMEKLGVDVAVIAEGMEAGGHIGKLTMTSLVRQVVEAVSIPVIAAGGIADGHGAAAA 180

Query: 181 FMLGADAVQVGRFVAKESNAHPNYKAKILKAKDIDTAVSAQVVGHPVRALKNKLVTY 240
 FMLGA+AVQ+GTRFVAKESNAH N+K KIL AKDIDT +SAQVVGHPVR++KNKL + Y
 Sbjct: 181 FMLGAEAVQIGTRFVAKESNAHQNFKDILAAKDIDTIVISAQVVGHPVRSIKNKLT SAY 240

Query: 241 SQA EKDYLAGRISINEIEELGAGALRNAVVDGDIVNGSV MAGQIAGLIKSEETCQEILED 300
 ++AEK +L G+ + +IEE+GAG+LR+AV++GDV+NGSV MAGQIAGL++ EE+C+ IL+D
 Sbjct: 241 AKA EKAFLLIGQKTATDIEEMGAGSLRHAVIEGDVVGNSVMAGQIAGLVRKEESCETILKD 300

Query: 301 IYSGARQVILSEASRWSDL 319
 IY GA +VI +EA RW +
 Sbjct: 301 IYYGAARVIQNEAKRWQSV 319

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1674

A DNA sequence (GBSx1778) was identified in *S.agalactiae* <SEQ ID 5193> which encodes the amino acid sequence <SEQ ID 5194>. This protein is predicted to be MCAT (fabD). Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1276(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with a *S.pneumoniae* sequence:

Identities = 203/306 (66%), Positives = 242/306 (78%), Gaps = 1/306 (0%)

Query: 1 MNKVSFLFAGQGAQKLGMDLYETFPPIVKETFDKASHVLGYDLRELIDKDLKLNQTKY 60
 M K +FLFAGQGAQ LGM RD Y+ +PIVKET D+AS VLG YDLR LID + DKLNQT+Y
 Sbjct: 1 MTKTAFLFAGQGAQYLGMDGRDFYDQYPIVKETIDRASQVLGYDLRYLIDTEEDKLNQTRY 60

-1884-

5 Query: 61 TQPAILTTSTAIYRLILKEIELRPMVAGLSLGEYSALVASGAIRFEDAVVLVARRGQLM 120
 TQPAIL TS AIYRL L+E +PDMVAGLSLGEYSALVASGA+ FEDAV LVA+RG M
 Sbjct: 61 TQPAILATSVAIYRL-LQEKGYQPDVAGLSLGEYSALVASGALDFEDAVALVAKRGAYM 119

10 Query: 121 EAAAPAGSGKMVAVLNADRQIIEACKKASQFGIVSPANYNTPKQIVIGGESIAVNAAVE 180
 E AAPA SGKMVAVLN ++IE+AC+KAS+ G+V+PANYNTP QIVI GE +AV+ AVE
 Sbjct: 120 EEAAPADSGKMVAVLNTPVEVIEEACQKASELGVVTPANYNTPAQIIVAGEVVAVDRAVE 179

15 Query: 241 IKSLLARQVMEPVRFDESIEETMKKMGMTQVVEIGPGKVLSGFLKKIDSSLSVHSVEDKIG 300
 I LL RQV EPVRF ESI M++ G++ +EIGPGKVLSGF+KKID + + VED+
 Sbjct: 240 IAQLLTRQVKEPVRFYESIGVMQEAGISNFIIEGPGKVLSGFVKKIDQTAHLAHVEDQAS 299

20 Query: 301 FNNLKE 306
 L E
 Sbjct: 300 LVALLE 305

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5195> which encodes the amino acid sequence <SEQ ID 5196>. Analysis of this protein sequence reveals the following:

25 Possible site: 19
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1602(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 201/299 (67%), Positives = 248/299 (82%), Gaps = 1/299 (0%)

35 Query: 1 MNKVSFLFAGQGAQKLGMDLYETFFPIVKETFDKASHVLGYDLRELIDKDLKLNQTKY 60
 M K +FLFAGQGAQKLGMD Y+ F IV++TFD+AS VLG YDLR LID D KLNQT Y
 Sbjct: 3 MTKTAFLLFAGQGAQKLGMDRYDNFAIVRKTFTDQASQVLGYDLRLIDSDELKLNQTSY 62

40 Query: 61 TQPAILTTSTAIYRLILKEIELRPMVAGLSLGEYSALVASGAIRFEDAVVLVARRGQLM 120
 TQPAILT+S AIYR +L ++PDMVAGLSLGEYSALVASGA+ FED + LVA+RG+LM
 Sbjct: 63 TQPAILTSSIAIYR-VLGLHHVKPDMVAGLSLGEYSALVASGALSFDLTLVAKRGRLM 121

45 Query: 121 EAAAPAGSGKMVAVLNADRQIIEACKKASQFGIVSPANYNTPKQIVIGGESIAVNAAVE 180
 E AAP GSGKMVA+V+N D Q+IE+ C+ A++ G+V+PANYNTP QIVIGG++ AVN AVE
 Sbjct: 122 EEAAPQGSGKMVAVMNTDVQVIEEVCQIAAKHGVVAPANYNTPSQIVIGGQTDVAVNAVE 181

50 Query: 181 ELKQQGVKRLIPLNVSGPFHTALLKPPASQKLSVDLKVHFSVSEIPVIGNTEAQIMKKDD 240
 LK++GVKRLIPLNVSGPFHTALL+PAS+ L+ L++ +FS +IP++GNTEA IM+KD
 Sbjct: 182 LLKERGVKRLIPLNVSGPFHTALLLEPASRLLAKELELYNFSDFKIPLVGNTEANIMEKDR 241

55 Query: 241 IKSLLARQVMEPVRFDESIEETMKKMGMTQVVEIGPGKVLSGFLKKIDSSLSVHSVEDKI 299
 I LLARQVMEPVRF +S+ T+ + G+TQ +E+GPGKVL+GF+KKID +L SVE+ +
 Sbjct: 242 IPELLARQVMEPVRFYDSVATLVESGITQFIEVGPGKVLTFGVKKIDKNLLCTSVENMV 300

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1675

A DNA sequence (GBSx1779) was identified in *S.agalactiae* <SEQ ID 5197> which encodes the amino acid sequence <SEQ ID 5198>. This protein is predicted to be beta-ketoacyl-ACP reductase (fabG).

60 Analysis of this protein sequence reveals the following:

-1885-

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.0930(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAF98275 GB:AF197933 beta-ketoacyl-ACP reductase [Streptococcus pneumoniae]
 Identities = 184/243 (75%), Positives = 212/243 (86%)

Query: 1 MQLKDKNIFITGSSRGIGLAIAHQFAQLGANIVLNGRSEISEDLIAEFADYGVKVIASG 60
 M+L+ KNIFITGSSRGIGLAIAH+FAQ GANIVLN R ISE+L+AEF++YG+KV+ ISG
 15 Sbjct: 1 MKLEHKNIFITGSSRGIGLAIAHKFAQAGANIVLNSRGAISEELIAEFSNYGIKVVPISG 60

Query: 61 DVSSFEDANRMIKEAIASLGSVDVLVNNAGITNDKMLKMTVEDFESVLKINLTGAFNMT 120
 DVS F DA RMI +AIA LGSVDVLVNNAGIT D LMLKMT DFE VLK+NLTGAFNMT
 20 Sbjct: 61 DVSDFADAKRMIDQAI AELGSVDVLVNNAGITQDTLMLKMT EADF EKVLKVNLTGAFNMT 120

Query: 121 QSVLKPMTKARQGAIINISSVVGLTGNVGQANYAASKAGLIGFTKSVAREVAARGIRVNA 180
 QSVLKPM KAR+GAIIN+SSVVGL GN+GQANYAASKAGLIGFTKSVAREVA+R IRVN
 25 Sbjct: 121 QSVLKPMKAREGAIINMSSVVGLMGNIGQANYAASKAGLIGFTKSVAREVASRNIRNV 180

Query: 181 IAPGFIESDMTDVIPEKMQEAILAQIPMKRIGKGKEVAQVASFLAEQEYLTGQVIAIDGG 240
 IAPG IESDMT ++ +K++EA LAQIPMK G+ ++VA + FLA Q+YLTGQV+ AIDGG
 30 Sbjct: 181 IAPGMIESDMTAILSDKIKEATLAQIPMK EFGQAEQVADLT VFLAGQDYLTGQVVAIDGG 240

Query: 241 MTM 243
 ++M
 30 Sbjct: 241 LSM 243

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3865> which encodes the amino acid sequence <SEQ ID 3866>. Analysis of this protein sequence reveals the following:

35 Possible site: 29
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.1088(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 201/244 (82%), Positives = 220/244 (89%)

45 Query: 1 MQLKDKNIFITGSSRGIGLAIAHQFAQLGANIVLNGRSEISEDLIAEFADYGVKVIASG 60
 M++K KNIFITGS+RGIGLA+AHQFA L ANIVLNGRS ISE+L+A F DYGV V+ ISG
 Sbjct: 1 MEIKGKNIFITGSTRGIGLAMAHQFASLEANIVLNGRSAISEELVASFTDYGVTVVTISG 60

50 Query: 61 DVSSFEDANRMIKEAIASLGSVDVLVNNAGITNDKMLKMTVEDFESVLKINLTGAFNMT 120
 DVS +A RM+ EAI SLGS+DVLVNNAGITNDKMLKMT ED FE VLKINLTGAFNMT
 Sbjct: 61 DVSEASEAKRMVN EAI ESLGSIDVLVNNAGITNDKMLKMT EEDFERVLKINLTGAFNMT 120

55 Query: 121 QSVLKPMTKARQGAIINISSVVGLTGNVGQANYAASKAGLIGFTKSVAREVAARGIRVNA 180
 QSVLKPM KARQGAIIN+SSVVGLTGN+GQANYAASKAG+IGFTKSVAREVAAR I VNA
 Sbjct: 121 QSVLKPMIKARQGAIINVSSVVGLTGNIGQANYAASKAGMIGFTKSVAREVAARNICVNA 180

60 Query: 181 IAPGFIESDMTDVIPEKMQEAILAQIPMKRIGKGKEVAQVASFLAEQEYLTGQVIAIDGG 240
 IAPGFIESDMT V+PEKMQE IL+QIPMKRIGK +EVA +ASFL EQ+Y+TGQVIAIDGG
 Sbjct: 181 IAPGFIESDMTGVLPEKMQEILSQIPMKRIGKAEVAHLASFLVEQDYITGQVIAIDGG 240

Query: 241 MTMQ 244
 MTMQ
 Sbjct: 241 MTMQ 244

-1886-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1676

- 5 A DNA sequence (GBSx1780) was identified in *S.agalactiae* <SEQ ID 5199> which encodes the amino acid sequence <SEQ ID 5200>. This protein is predicted to be 3-oxoacyl-(acyl-carrier-protein) synthase II (fabF). Analysis of this protein sequence reveals the following:

```

Possible site: 51
>>> Seems to have no N-terminal signal sequence
10    INTEGRAL    Likelihood = -0.37    Transmembrane    338 - 354 ( 338 - 354)

----- Final Results -----
                bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF98276 GB:AF197933 beta-ketoacyl-ACP synthase II
[Streptococcus pneumoniae]
20    Identities = 340/410 (82%), Positives = 375/410 (90%)

Query: 1    MTLQRVVVTGYGVTSPIGNTPEEFWNSLKEGNGVIGIPITKFDSSDFMVKNAAEIHDFFPD 60
          M L RVVVTGYGVTSPIGNTPEEFWNSL G +GIG ITKFD SDF V NAAEI DFFPD
25    Sbjct: 1    MKLNRVVVTGYGVTSPIGNTPEEFWNSLATGKIGIGGITKFDHSDFDVHNAAEIQDFFPD 60

Query: 61    KYFVKKDLNRFDMYSLYALYASSEAIQHANLNLDEIDADRFGVIVASGIGGIQEIEEQVI 120
          KYFVKKD NRPD YSLYALYA+ EA+ HANL+++ ++ DRFGVIVASGIGGI+EIE+QV+
30    Sbjct: 61    KYFVKKDITNRFDMYSLYALYAAQEAVNHANLDVEALNRDRFGVIVASGIGGIKEIEDQVL 120

Query: 121   RLHEKGPKRVKPMTLPKALPNMAAGNVAMRLGAHGVCKSINTACASSNDAIGDAFRNIKF 180
          RLHEKGPKRVKPMTLPKALPNMA+GNVAMR GA+GVCKSINTAC+SSNDAIGDAFR+IKF
35    Sbjct: 121  RLHEKGPKRVKPMTLPKALPNMASGNVAMRFGANGVCKSINTACSSNDAIGDAFRSIKF 180

Query: 181   GIQDIMVVGGAEEAITKFAIAGFQSLTALSTTEDPSRASIPFDKDRNGFIMGEGSGMLVL 240
          G QD+M+VGG EA+IT FAIAGFQ+LTALSTTEDP+RASIPFDKDRNGF+MGEKSGMLVL
40    Sbjct: 181  GFQDVMLVGGTEASITPFAIAGFQALTALSTTEDPTRASIPFDKDRNGFVMGEGSGMLVL 240

Query: 241   ESLEHAEKRGATILAEVVGYGNTCDAYHMTSPHPEGLGATKAIQLALVEANIKPEEVNVV 300
          ESLEHAEKRGATILAEVVGYGNTCDAYHMTSPHPEG GA KAI+LAL EA I PE+V YV
45    Sbjct: 241  ESLEHAEKRGATILAEVVGYGNTCDAYHMTSPHPEGQAIKAIKLALAEAEISPEQVAVV 300

Query: 301   NAHGTSTPANKEGESQAIVAALGTDVPVSSTKSFTGHLLGAAGAVEAIAITIEAIRHSYVP 360
          NAHGTSTPANKEGES AIVA LG +VPVSSTKSFTGHLLGAAGAVEAI TIEA+RH++VP
50    Sbjct: 301  NAHGTSTPANKEGESGAIVAVLGKEVPVSSTKSFTGHLLGAAGAVEAIVTIEAMRHNFPV 360

Query: 361   MTAGTTELSEDITANVIFGQGQDADIRYAIISNTFGFGGHNAVLAFKRWED 410
          MTAGT+E+S+ I ANV++GQG + +I YAIISNTFGFGGHNAVLAFKRWE+
55    Sbjct: 361  MTAGTSEVSDYIEANVVGQGLEKIPYAIISNTFGFGGHNAVLAFKRWEN 410

```

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3851> which encodes the amino acid sequence <SEQ ID 3852>. Analysis of this protein sequence reveals the following:

```

Possible site: 51
>>> Seems to have no N-terminal signal sequence
55    ----- Final Results -----
                bacterial cytoplasm --- Certainty=0.0890(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```


-1887-

An alignment of the GAS and GBS proteins is shown below.

Identities = 346/410 (84%), Positives = 377/410 (91%)

```

5  Query: 1  MTLQRVVVTGYGVTSPIGNTPEEFWNSLKEGNVGIGIPITKFDSSDFMVKNAAEIHDFPFD 60
    MT +RVVVVTGYG+TSPIG+ PE FWN+LK G +GIGIPITKFD++D+ VKNAAEI DFPFD
    Sbjct: 1  MTFKRVVVTGYGLTSPIGHDPETFWNNLKAGQIGIGIPITKFDITDYAVKNAAEIQDFPFD 60

    Query: 61  KYFVKDLNRFDMYSLYALYASSEAIQHANLNLDEIDADRFGVIVASGIGGIQEIEEQVI 120
    KYFVKDLNRFD YSLYALYA+ EAI HA+LN++ +D+DRFGVIVASGIGGI EIEEQVI
10  Sbjct: 61  KYFVKDLNRFDRYSLYALYAAKEAINHADLNIEMVDSRFGVIVASGIGGIAEIEEQVI 120

    Query: 121  RLHEKGPKRVKPMTLPKALPNMAAGNVAMRLGAHGVCKSINTACASSNDAIGDAFRNIKF 180
    RLHEKGPKRVKPMTLPKALPNMAAGNVAM L A GVCKSINTACASSNDAIGDAFR IKF
15  Sbjct: 121  RLHEKGPKRVKPMTLPKALPNMAAGNVAMSLKAQGVCKSINTACASSNDAIGDAFRAIKF 180

    Query: 181  GIQDIMVVGAEAAITKFAIAGFQSLTALSTTEDPSRASIPFDKDRNGFIMGEGSGMLVL 240
    G QD+M+VGG+EAATKFAIAGFQSLTALSTTEDPSR+SIPFDKDRNGFIMGEGSGMLVL
    Sbjct: 181  GTQDMIVVGSEAAITKFAIAGFQSLTALSTTEDPSRSSIPFDKDRNGFIMGEGSGMLVL 240

20  Query: 241  ESLEHAERKATILAEVVGYNCTDAYHMTSPHPEGLGATKAIQLALVEANIKPEEVNYV 300
    ESLEHA++RGATILAE+VGYGNTCDAYHMTSP+PEGLGA KAI LAL EA I+ +NYV
    Sbjct: 241  ESLEHAQERKATILAEIVGYGNTCDAYHMTSPNPEGLGARKAIHLALQEAGIEASAINV 300

25  Query: 301  NAHGTSTPANKEGESQAIVAALGTDVPVSSTKSFTGHLLGAAGAVEAIATIEAIRHSYVP 360
    NAHGTSTPANKEGESQAIVA LG DVPVSSTKSFTGHLLGAAGA+EAIATIEA+RH+YVP
    Sbjct: 301  NAHGTSTPANKEGESQAIVAVLGKDVPSSTKSFTGHLLGAAGAIEAIATIEAMRHNYVP 360

    Query: 361  MTAGTTELSEDITANVIFGQSQDADIRYAISNTFGFGGHNAVLAFKRWED 410
    MTAGT LSEDI ANVIFG+G++ I YAISNTFGFGGHNAVLAFK WE+
30  Sbjct: 361  MTAGTQALSEDIEANVIFGEGKETAINYIAISNTFGFGGHNAVLAFKCWEE 410

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1677

35 A DNA sequence (GBSx1781) was identified in *S.agalactiae* <SEQ ID 5201> which encodes the amino acid sequence <SEQ ID 5202>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

```

    bacterial cytoplasm --- Certainty=0.3052(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 A related GBS nucleic acid sequence <SEQ ID 9553> which encodes amino acid sequence <SEQ ID 9554> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF98277 GB:AF197933 biotin carboxyl carrier protein
[Streptococcus pneumoniae]

50 Identities = 103/169 (60%), Positives = 127/169 (74%), Gaps = 11/169 (6%)

```

    Query: 19  LDIQEIKDLMTQFDESSLREFSFKTSDGELSFSKNEGKAPLVPTMSPMSHQPEATPTTIAT 78
    +++ +IKDLMTQFD+SSLREFS+K EL FSKNE + VP ++ Q P +AT
    Sbjct: 1  MNLNDIKDLMTQFDQSSLREFSYKNGTDELQFSKNEARP--VPEVAT---QVAPAPVLAT 55

```

```

55  Query: 79  PVSNEAGEQTKQATEVVSEIP---ESTVTVAEGDVVESPLVGVAYLASGPDKNFVSVGD 135
    P + + A V E+P E++V EG++VESPLVGV YLA+GPDKP FV+VGD
    Sbjct: 56  P--SPVAPTSAPAETVAEEVPAPAEASVAT-EGNLVESPLVGVVYLAAGPDKPAFVTVGD 112

```

-1888-

Query: 136 SVKKGQTLMIIEAMKVMNEVPAPHDGVVTEILVANEVIEFGKGLVRIK 184
 SVKKGQTL+IIEAMKVMNE+PAP DGVVTEILV+NEE++EFGKGLVRIK
 Sbjct: 113 SVKKGQTLVIIEAMKVMNEIPAPKDGVVTEILVSNEEMVEFGKGLVRIK 161

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5203> which encodes the amino acid sequence <SEQ ID 5204>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3132(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 107/171 (62%), Positives = 126/171 (73%), Gaps = 10/171 (5%)

Query: 19 LDIQEIKDLMTQFDESSLREFSFKTSDGELSFSGKNEGKAPLVPTMSPMSHQPEATPT--- 75
 L+IQEIKDLM QFD SSLREF FKT++GEL FSKNE + S+Q A P
 20 Sbjct: 1 LNTIQEIKDLMAQFDTSSLREFLFKTNEGELIFSKNEQHLN-----ASTSNQEHAVPVPQV 55
 Query: 76 --IATPVSNEAGEQTKQATEVVSSEIPESTVTVVAEGDVVESPLVGVAYLASGPDKPNFVSV 133
 + P ++EA V E P++ VAEGD+VESPLVGVAYLA+ PDKP FV+V
 Sbjct: 56 QLVNPFTASEASSPASVKDVPVEEQPQAESFVAEGDIVESPLVGVAYLAASPDKPPFVAV 115
 25 Query: 134 GDSVKKGQTLMIIEAMKVMNEVPAPHDGVVTEILVANEVIEFGKGLVRIK 184
 GD+VKKGQTL+IIEAMKVMNEVPAP DGV+TEILV+NE+VIEFG+GLVRIK
 Sbjct: 116 GDTVKKGQTLVIIEAMKVMNEVPAPCDGVITEILVSNEEDVIEFGQGLVRIK 166

- 30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1678

A DNA sequence (GBSx1782) was identified in *S.agalactiae* <SEQ ID 5205> which encodes the amino acid sequence <SEQ ID 5206>. This protein is predicted to be beta-hydroxyacyl-ACP dehydratase (fabZ).

- 35 Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2267(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:AAF98278 GB:AF197933 beta-hydroxyacyl-ACP dehydratase
 [Streptococcus pneumoniae]
 Identities = 130/140 (92%), Positives = 135/140 (95%)
 Query: 1 MIDIKEIREALPHRYPMLLVDRVLEVSEDEIVAIAKNVSINEPFFNGHFPEYPVMPGVLM 60
 MIDI+ I+EALPHRYPMLLVDRVLEVSED IVAIAKNV+INEPFFNGHF+YPVMPGV+IM
 50 Sbjct: 1 MIDIQGIKEALPHRYPMLLVDRVLEVSEDTIVAIAKNVTINEPFFNGHFQYPVMPGVVIM 60
 Query: 61 EALAQTAGVLELSKEENKGLVFYAGMDKVKFKKQVVPDQLVMTAKFVKRRGTIAVVEA 120
 EALAQTAGVLELSK ENKGLVFYAGMDKVKFKKQVVPDQLVMTA FVKRRGTIAVVEA
 55 Sbjct: 61 EALAQTAGVLELSKPENKGLVFYAGMDKVKFKKQVVPDQLVMTATFVKRRGTIAVVEA 120
 Query: 121 IAEVDGKLAASGTLTFAIGN 140
 AEVDGKLAASGTLTFAIGN

-1889-

Sbjct: 121 KAEVDGKLAASGTLTFAIGN 140

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5207> which encodes the amino acid sequence <SEQ ID 5208>. Analysis of this protein sequence reveals the following:

5 Possible site: 59
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.1882(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 127/139 (91%), Positives = 133/139 (95%)

Query: 1 MIDIKEIREALPHRYPMLLVDRVLEVSDEIVAIAKNSINEPFFNGHFPPEYPMVPGVLIM 60
 M+DI+EI+ ALPHRYPMLLVDRVLEVS+D IVAIKNV+INEPFFNGHFP YPVMVPGVLIM
 Sbjct: 1 MMDIREIQALPHRYPMLLVDRVLEVSDDHIVAIAKNVTINEPFFNGHFPHPYPMVPGVLIM 60

20 Query: 61 EALAQTAGVLELSKEENKGKLVFYAGMDKVKFKKQVVPDQLVMTAKFVKRRGTIAVVEA 120
 EALAQTAGVLELSKEENKGKLVFYAGMDKVKFKKQVVPDQLVMTA F+KRRGTIAVVEA
 Sbjct: 61 EALAQTAGVLELSKEENKGKLVFYAGMDKVKFKKQVVPDQLVMTATFIKRRGTIAVVEA 120

25 Query: 121 IAEVDGKLAASGTLTFAIG 139
 AEVDGKLAASGTLTFA G
 Sbjct: 121 RAEVDGKLAASGTLTFACG 139

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 1679

A DNA sequence (GBSx1783) was identified in *S.agalactiae* <SEQ ID 5209> which encodes the amino acid sequence <SEQ ID 5210>. This protein is predicted to be acetyl-coenzyme A carboxylase, biotin carboxylase (accC). Analysis of this protein sequence reveals the following:

35 Possible site: 30
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 40 bacterial cytoplasm --- Certainty=0.1203(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF98279 GB:AF197933 acetyl-CoA carboxylase biotin carboxylase
 subunit [Streptococcus pneumoniae]
 45 Identities = 361/451 (80%), Positives = 405/451 (89%)

Query: 1 MFKKILIANRGEIAVRIIRAAREMGISTVAIYSEADKESLHTILADEAICVGPAKSAESY 60
 MF+KILIANRGEIAVRIIRAARE+GI+TVA+YS ADKE+LHT+LADEA+C+GP K+ ESY
 Sbjct: 1 MFRKILIANRGEIAVRIIRAARELGATVAVYSTADKEALHTLLADEAIVCIGPGKATESY 60

50 Query: 61 LNVNAILSAAIVTGAEAVHPGFGFLSENSKFATMCEEMNLKFIGPSGEVMDKMGDKINAR 120
 LN+NA+LSAA++T AEA+HPGFGFLSENSKFATMCEE+ +KFIGPSG VMD MGDKINAR
 Sbjct: 61 LNINAVLSAAVLTEAEAIHPGFGFLSENSKFATMCEEVGKIFIGPSGHVMDMMGDKINAR 120

55 Query: 121 TEMIKADVPIPGSDGQVTSVEEAVSIAEEIGYPLMLKASAGGGGKGIRKVKSADELKPA 180
 +MIKA VPVPIPGSDG+V + EEA+ +AE+IGYP+MLKASAGGGGKGIRKV+ D+L A
 Sbjct: 121 AQMIKAGVPVPIPGSDGEVHNSEELIVAEEKIGYPVMLKASAGGGGKGIRKVEKPDDLVA 180

-1890-

Query: 181 FESASQEALAAFGNGAMYIEKVIYPARHIEVQILGDSFGKIVHLGERDCSLQRNNQKVLE 240
 FE+AS EA A +GNGAMYIE+VIYPARHIEVQILGD G ++HLGERDCSLQRNNQKVLE
 Sbjct: 181 FETASSEAKANYGNGAMYIERVIYPARHIEVQILGDEHGHVHHLGERDCSLQRNNQKVLE 240

5 Query: 241 ESPSVAIGNTLRQQIGEEAAVRAAEAVSYENAGTIEFLLDENSGQFYFMMNTRVQVEHPV 300
 ESPS+AIG TLR +IG AAVRAAE V YENAGTIEFLLD E S FYFMMNTRVQVEHPV
 Sbjct: 241 ESPSIAIGKTLRHEIGAAAVRAAEFVGYENAGTIEFLLD EASNIFYFMMNTRVQVEHPV 300

10 Query: 301 TEFVTGVDIVKEQIRIAAGIPLSVS QNDIKLTGHAIECRINAENPQNFAPCPGTINGLH 360
 TEFV+GVDIVKEQI IAAG PLSV Q DI L GHAEICRINAENP FNFAP PG I L+
 Sbjct: 301 TEFVSGVDIVKEQICIAAGQPLSVKQEDIVLRGHAIECRINAENPAFNFAPSPGKITNLY 360

15 Query: 361 LPAGGMGLRVDSAVYTGTYTIPPYYDSMIKIVHGENRFDALMKMQRALYELEIDGIVTN 420
 LP+GG+GLRVDSAVY GYTIPPYYDSMIK+IVHGENRFDALMKMQRALYELEI+G+ TN
 Sbjct: 361 LPSSGGVGLRVDSAVYPGYTIPPYYDSMIKIVHGENRFDALMKMQRALYELEIEGVQTN 420

20 Query: 421 TEFQMDLISDKKVLADYDTSFLMEDFLPRY 451
 +FQ+DLISD+ V+AGDYDTSFLME FLP+Y
 Sbjct: 421 ADFQLDLISDRNVIAGDYDTSFLMETFLPKY 451

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5211> which encodes the amino acid sequence <SEQ ID 5212>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1784(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 369/451 (81%), Positives = 421/451 (92%)

Query: 1 MFKKILIANRGEIAVRIIRAAREMGISTVAIYSEADKESLHTILADEAICVGPAKSAESY 60
 MFKKILIANRGEIAVRIIRAARE+GISTVA+YSEADKE+LHTILADEAIC+GPA+S ESY
 Sbjct: 17 MFKKILIANRGEIAVRIIRAARELGISTVAVYSEADKEALHTILADEAICIGPARSKESY 76

Query: 61 LNVNAILSAAIVTGAEAVHPGFGFLSENSKFATMCEEMNLKFIGPSGEVMDKMGDKINAR 120
 LN+N++LSAAIVTGA+A+HPGFGFLSENSKFATMCEEMN+KFIGPS VMDKMGDKINAR
 Sbjct: 77 LNMNSVLSAAIVTGAQAIHPGFGFLSENSKFATMCEEMNIKFIGPSASVMDKMGDKINAR 136

Query: 121 TEMIKADVPVPIPGSDGQVTSVEEAVSIAEEIGYPLMLKASAGGGGKGIRKVSADLEKPA 180
 +EMIKA VPVIPGSDG+V + +EA++IA +IGYP+MLKASAGGGGKGIRKV++ +L+ A
 Sbjct: 137 SEMIKAGVPVPIPGSDGEVYNAQALAIANKIGYPVMLKASAGGGGKGIRKVETADLEAA 196

Query: 181 FESASQEALAAFGNGAMYIEKVIYPARHIEVQILGDSFGKIVHLGERDCSLQRNNQKVLE 240
 F +ASQEAL AFGNGAMY+EKVIYPARHIEVQILGD++G I+HLGERDCSLQRNNQKVLE
 Sbjct: 197 FNAASQEALGAFNGAMYLEKVIYPARHIEVQILGDAYGNIHHLGERDCSLQRNNQKVLE 256

Query: 241 ESPSVAIGNTLRQQIGEEAAVRAAEAVSYENAGTIEFLLDENSGQFYFMMNTRVQVEHPV 300
 ESPS+AIGNTLR ++G+AAVRAAEAV+YENAGTIEFLLD E+S +FYFMMNTR+QVEHPV
 Sbjct: 257 ESPSIAIGNTLRHEMGQAAVRAAEAVAYENAGTIEFLLD EDESEKIFYFMMNTRI QVEHPV 316

Query: 301 TEFVTGVDIVKEQIRIAAGIPLSVS QNDIKLTGHAIECRINAENPQNFAPCPGTINGLH 360
 TEFVTGVDIVKEQI+IAAG PL+++Q DI +TGHAEICRINAEN FNFAP PG I L+
 Sbjct: 317 TEFVTGVDIVKEQIKIAAGQPLAINQEDITITGHAEICRINAENTAFNFAPSPGKITDLY 376

Query: 361 LPAGGMGLRVDSAVYTGTYTIPPYYDSMIKIVHGENRFDALMKMQRALYELEIDGIVTN 420
 +P+GG+GLRVDSAVY GY IPPYYDSMIK+IVHG NRFDALMKMQRAL ELEI+GI+TN
 Sbjct: 377 MPSSGGVGLRVDSAVYNGYAIPPPYYDSMIKIVHGSNRFDALMKMQRALVELEIEGIITN 436

Query: 421 TEFQMDLISDKKVLADYDTSFLMEDFLPRY 451
 T+FQ+DLISDK+V+AGDYDTSFLME FLP Y
 Sbjct: 437 TDFQLDLISDKRVIAGDYDTSFLMETFLPHY 467

-1891-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1680

A DNA sequence (GBSx1784) was identified in *S.agalactiae* <SEQ ID 5213> which encodes the amino acid sequence <SEQ ID 5214>. This protein is predicted to be acetyl-CoA carboxylase beta subunit (accD). Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3571(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF98280 GB:AF197933 acetyl-CoA carboxylase beta subunit
[Streptococcus pneumoniae]
Identities = 221/285 (77%), Positives = 248/285 (86%), Gaps = 1/285 (0%)

20 Query: 1 MALFSKKDKYIRISPNKALGSSDKRSLPEVPDELFAKCPCKHMIYQKDLGLAKICPACS 60
MALFSKKDKYIRI+PN+++ + PEVPDELFP++CP CKH IYQKDLG +ICP CS
Sbjct: 1 MALFSKKDKYIRINPNRSVREKPOAK-PEVPDELFSQCPGCKHTIYQKDLGSRICPHCS 59

25 Query: 61 YNFRISAQERLLLLTVDEDSFEELFTGIETKDPLNFPNYREKLAATRQKTNLDEAVVTGLA 120
Y FRISAQERL LT+D +F+ELFTGIE+KDPL+FP Y++KLA+ R+KT L EAVVTG A
Sbjct: 60 YTFRISAQERLALTIDMGTFKELFTGIESKDPLHFPYQKKLASMREKTGLHEAVVTGTA 119

30 Query: 121 KIKGQTTALAIMDSHFIMASMGTVVGEKLTRLFELATEKKLPVIFTASGGARMQEGIMS 180
IKGQT AL IMDS+FIMASMGTVVGEK+TRLFE AT +KLP+V+FTASGGARMQEGIMS
Sbjct: 120 LIKGQTVLALGIMDSNFIMASMGTVVGEKITRLFEYATVEKLPVVLFTASGGARMQEGIMS 179

35 Query: 181 LMQMAKVSAAVKRHSNQGLFYLTILTDPPTGGVTASFAMEGDIILAEPQALVGFAGRRVI 240
LMQMAK+SAAVKRHSN GLFYLTILTDPPTGGVTASFAMEGDIILAEPQ+LVGFAGRRVI
Sbjct: 180 LMQMAKISAAVKRHSNAGLFYLTILTDPPTGGVTASFAMEGDIILAEPQSLVGFAGRRVI 239

40 Query: 241 ETTVREDLPEGFQKAEFLLEHGFVDALINRTELRCIAQLIAFHG 285
E TVRE LPE FQKAEFLLEHGFVDAL+ R +L D IA L+ HG
Sbjct: 240 ENTVRESLPEDFQKAEFLLEHGFVDALVRRDLPTIASLVRLHG 284

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5215> which encodes the amino acid sequence <SEQ ID 5216>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4092(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 232/285 (81%), Positives = 253/285 (88%)

55 Query: 1 MALFSKKDKYIRISPNKALGSSDKRSLPEVPDELFAKCPCKHMIYQKDLGLAKICPACS 60
MALF KKDKYIRI+PN +L S ++PEVPDELFAKCP+CKHMIY+KDLGLAKICP CS
Sbjct: 1 MALFRKKDKYIRITPNNSLKGSVSHNVPEVPDELFAKCPCKHMIYKKDLGLAKICPTCS 60

Query: 61 YNFRISAQERLLLLTVDEDSFEELFTGIETKDPLNFPNYREKLAATRQKTNLDEAVVTGLA 120
YNFRISAQERL LTVDE SF+ELFT IETKDPL FP Y+EKL ++ T L EAV+TG A

-1892-

Sbjct: 61 YNFRISAQERLTLTVDEGSFQELFTSIETKDP LRFPGYQEK LQKAKETTGLHEAVLTGKA 120

Query: 121 KIKGQTTALAIMDSHFIMASMGTVVGEK LTRLFELATEKKLPVIFTASGGARMQEGIMS 180
+K Q ALAIMDSHFIMASMGTVVGEK+TRLFELA E+ LP+VIFTASGGARMQEGIMS

5 Sbjct: 121 MVKEQKIALAIMDSHFIMASMGTVVGEKITR LFE LAIEENLPV VIFTASGGARMQEGIMS 180

Query: 181 LMQMAKVSAAVKRHSNQLFYLTILTDP TTGGVTASFAMEGDIILAEPQALVGFAGRRVI 240
LMQMAKVSAAVKRHSN GLFYLTILTDP TTGGVTASFAMEGDIILAEPQ+LVGFAGRRVI

10 Sbjct: 181 LMQMAKVSAAVKRHSNAGLFYLTILTDP TTGGVTASFAMEGDIILAEPQSLVGFAGRRVI 240

Query: 241 ETTVREDLPEGFQKAEFLLEHGFVD A IINRTEL RDCIAQLIAFHG 285
ETTVRE+LP+ FQKAEFL +HGFVD A I+ RTEL R D I A L+AFHG

15 Sbjct: 241 ETTVRENLPDDFQKAEFLQDHGFVD A IVKRTEL RDKIAHLVAFHG 285

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1681

20 A DNA sequence (GBSx1785) was identified in *S. agalactiae* <SEQ ID 5217> which encodes the amino acid sequence <SEQ ID 5218>. This protein is predicted to be acetyl-CoA carboxylase alpha subunit (accA). Analysis of this protein sequence reveals the following:

Possible site: 50
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.22 Transmembrane 149 - 165 (149 - 165)

25 ----- Final Results -----
bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30 A related GBS nucleic acid sequence <SEQ ID 9555> which encodes amino acid sequence <SEQ ID 9556> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF98281 GB:AF197933 acetyl-CoA carboxylase alpha subunit
[Streptococcus pneumoniae]

35 Identities = 186/254 (73%), Positives = 222/254 (87%)

Query: 13 DVTRILKDARDQGR L TALDYAELIFDNFMELHGD RQFADDKSIIGGLGYLAGRPVTIVGI 72
++ +I+++AR+Q RLT LD+A IFD F++LHGDR F DD +++GG+G+L + VT+VGI

40 Sbjct: 2 NIAKIVREAREQSRLTTLDFATGIFDEFIQ LHGDRSFRDDGAVVGGIGWLGDAQVTVVGI 61

Query: 73 QKGKNLQDNLD RHFQPHPEGYRKALRLMKQAEKFGRPVITFINTAGAYPGVGAEERGQG 132
QKGK+LQDNL R+FGQPHPEGYRKALRLMKQAEKFGRPV+TFINTAGAYPGVGAEERGQG

45 Sbjct: 62 QKGKSLQDNLKRNFQPHPEGYRKALRLMKQAEKFGRPVVT FINTAGAYPGVGAEERGQG 121

Query: 133 EAIARNLLEMSDLKVPIIAIIIGEGSGGALALAVADKVMLEHTVYSILSPEGFASILW 192
EAIARNL+EMSDLKVPIIAIIIGEGSGGALALAVAD+VWMLE+++Y+ILSPEGFASILW

50 Sbjct: 122 EAIARNLMEMSDLKVPIIAIIIGEGSGGALALAVADRVWMLENSIYAILSPG FASILW 181

Query: 193 KDGT RTTEAAQLMKMTAGELYHMEVVDKVIPEHGYFSSEIVDMIKTSLISELEVLSQLSL 252
KDGT R EAA+LMK+T+ EL M+VVDKVI E G S E++ +K L +EL LSQ L

55 Sbjct: 182 KDGT RAMEAAELMKITSHELLEMDVVDKVI SEIGLSSKELIKSVKKE LQTELARLSQKPL 241

Query: 253 EDLLEQRYQRF RKY 266
E+LLE+RYQRF RKY

Sbjct: 242 EELLEERYQRF RKY 255

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 5219> which encodes the amino acid sequence <SEQ ID 5220>. Analysis of this protein sequence reveals the following:

-1893-

Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.22 Transmembrane 139 - 155 (139 - 155)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:AAF98281 GB:AF197933 acetyl-CoA carboxylase alpha subunit
 [Streptococcus pneumoniae]
 Identities = 189/254 (74%), Positives = 225/254 (88%)

15 Query: 3 DVSRILKEARDQGRLTTLDYANLIFDDFMELHGDRHFSDDGAIVGGLAYLAGQPVTIVIGI 62
 ++++I++EAR+Q RLTTLD+A IFD+F++LHGDR F DDGA+VGG+ +L Q VTV+GI
 Sbjct: 2 NIAKIVREAREQSRLTTLDATGIFDEFIQLHGDRSFRDDGAVVGGIGWLGDQAVTVVIGI 61

20 Query: 63 QKGKNLQDNLARNFGQPNPEGYRKALRLMKQAEKFGRPVVTFFINTAGAYPGVGAEERGQG 122
 QKGK+LQDNL RNFGQP+PEGYRKALRLMKQAEKFGRPVVTFFINTAGAYPGVGAEERGQG
 Sbjct: 62 QKGKSLQDNLKRNFGQPHPEGYRKALRLMKQAEKFGRPVVTFFINTAGAYPGVGAEERGQG 121

25 Query: 123 EAIAKNLMEMSDLKVPITIIIIIGEGSGGALALAVADQVWMLENTMYAVLSPEGFASILW 182
 EAIA+NL+MEMSDLKVPITIIIIIGEGSGGALALAVAD+VWMLENT+YA+LSPEGFASILW
 Sbjct: 122 EAIARNLMEMSDLKVPITIIIIIGEGSGGALALAVADRVWMLENTSIYAILSPGFASILW 181

30 Query: 183 KDGSRATEAAELMKITAGELYKMGIVDRIIPEHGYFSSEIVDIIKANLIEQITSLQAKPL 242
 KDG+RA EAAELMKIT+ EL +M +VD++I E G' S E++ +K L ++ L KPL
 Sbjct: 182 KDGTRAMEAAELMKITSHELLEMDVVDKVISEIGLSSKELIKSVKKELQTELARLSQKPL 241

Query: 243 DQLLDERYQRFKY 256
 ++LL+ERYQRFKY
 Sbjct: 242 EELLEERYQRFKY 255

35 An alignment of the GAS and GBS proteins is shown below.

Identities = 204/254 (80%), Positives = 236/254 (92%)

40 Query: 13 DVTRILKDARDQGRLTALDYAELIFDNFMELHGDRQFADDKSIIGGLYLAGRPVTIVIGI 72
 DV+RILK+ARDQGRLT LDYA LIFD+FMELHGDR F+DD +I+GGL YLAG+PVT++GI
 Sbjct: 3 DVSRILKEARDQGRLTTLDYANLIFDDFMELHGDRHFSDDGAIVGGLAYLAGQPVTIVIGI 62

45 Query: 73 QKGKNLQDNLDRHFGQPHPEGYRKALRLMKQAEKFGRPVITFFINTAGAYPGVGAEERGQG 132
 QKGKNLQDNL R+FGQP+PEGYRKALRLMKQAEKFGRPV+TFINTAGAYPGVGAEERGQG
 Sbjct: 63 QKGKNLQDNLARNFGQPNPEGYRKALRLMKQAEKFGRPVVTFFINTAGAYPGVGAEERGQG 122

50 Query: 133 EAIARNLLEMSDLKVPITIIIIIGEGSGGALALAVADKVMLEHTVYSIISPEGFASILW 192
 EAIA+NL+EMSDLKVPITIIIIIGEGSGGALALAVAD+VWMLE+T+Y++LSPEGFASILW
 Sbjct: 123 EAIAKNLMEMSDLKVPITIIIIIGEGSGGALALAVADQVWMLENTMYAVLSPEGFASILW 182

55 Query: 193 KDGTRTTEAAQLMKMTAGELYHMEVVDKVIPEHGYFSSEIVDMIKTSLISELEVLSQLSL 252
 KDG+R TEAA+LMK+TAGELY M +VD++IPEHGYFSSEIVD+IK +LI ++ L L
 Sbjct: 183 KDGSRATEAAELMKITAGELYKMGIVDRIIPEHGYFSSEIVDIIKANLIEQITSLQAKPL 242

Query: 253 EDLLEQRYQRFKY 266
 + LL++RYQRFKY
 Sbjct: 243 DQLLDERYQRFKY 256

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1894-

Example 1682

A DNA sequence (GBSx1786) was identified in *S.agalactiae* <SEQ ID 5221> which encodes the amino acid sequence <SEQ ID 5222>. This protein is predicted to be sakacin A production response regulator. Analysis of this protein sequence reveals the following:

```

5      Possible site: 56
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3304(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9557> which encodes amino acid sequence <SEQ ID 9558> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

      >GP:BAA88824 GB:AB016077 sakacin A production response regulator
      [Streptococcus mutans]
      Identities = 76/142 (53%), Positives = 99/142 (69%)

20      Query: 36  MQTFKAKGQLARNSFTELSRALEQRMDFKMQRVSNWANQAQVGRPHFWVYYRKDTDQLD 95
      M K GQ AR FTE+++ L ++ F+M RVSNWANQAQV RPHFW YY++ D D
      Sbjct: 1  MIALKTLGQSARAEFTEIAKVLALKVSPFEMMRVSNWANQAQVVRPHFWCYKQPEDNQD 60

      Query: 96  DVAAVALRVYGVKDSFGVSLVSVFVERQKSDKTLEKQARVLSIPTASPLYFMVQRQGETHR 155
      DV +A+R+YG +FG+S+EVSF+ER+KS TL KQ +VL IPIA PLY+ Q + E+HR
25      Sbjct: 61  DVGLAIRLYGNSANFGISVEVSFIERKKSKATLAKQHKVLDIPIAEPLYFFAQEKSESHR 120

      Query: 156  EEGNEENRQRLMQEIKSGKVRK 177
      G E RQ L Q++ G+VRK
30      Sbjct: 121 VSGTEAYRQMLRQKVADGQVRK 142

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1683

A DNA sequence (GBSx1787) was identified in *S.agalactiae* <SEQ ID 5223> which encodes the amino acid sequence <SEQ ID 5224>. This protein is predicted to be seryl-tRNA synthetase (serS). Analysis of this protein sequence reveals the following:

```

      Possible site: 60
40      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1866(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

      >GP:CAB11789 GB:Z99104 seryl-tRNA synthetase [Bacillus subtilis]
      Identities = 262/425 (61%), Positives = 322/425 (75%), Gaps = 1/425 (0%)

50      Query: 1  MLDLKRIRTDFFDVVAKKLATRGVDQETLTLTKELDIKRRELLIKAEFAKAQRNVASAAIA 60
      MLD K +R +F + KL +G D + LD +RREL+ K EE K +RN S +A
      Sbjct: 1  MLDTKMLRANFQEIKAQLVHKGEDLTDQDKFEALDDRRRELIGKVEELKGKRNEVSQQVA 60

```


-1895-

Query: 61. QAKRNKENADEQIAAMQTLASADIKADAELADVANDLQSMVTVLPNTPADDPVPLGADEDE 120
 KR K++AD I M+ + +IK +D EL V+A L +++ +PN P + VP+G ED+
 Sbjct: 61 VLKREKKDADHIIKEMREVGEIEIKLDEELRTVEAELEDTILLISIPNIPHESVPVGETEDD 120

5 Query: 121 NVEVRRWGTTPREFDFETKAHWDLGESLGILDWERGAKVTGSRFLFYKGLGARLERAIYSF 180
 NVEVR+WG F +E K HWD+ + LGILD+ER AKVTGSRF+FYKGLGARLERAIY+F
 Sbjct: 121 NVEVRKWEKPSFAYEPKPHWDIADLGLDIFERA AKVTGSRFVYKGLGARLERALYNF 180

10 Query: 181 MLDEHAKE-GYTEVIPPYVMVNHDSMFGTGQYPKFKEDTFELADSPFVLIPTAEVPLTNY 239
 MLD H E YTEVIPPYVMV N SM GTGQ PKF+ED F++ + + LIPTAEVP+TN +
 Sbjct: 181 MLDLHVDEYNYTEVIPPYVMVNRASMTGTGQLPKFEEDAFKIREEDYFLIPTAEVPITNMH 240

15 Query: 240 RDEIIDGKELPIYFTAMSPSFRSEAGSAGRDTRGLIRLHQFHKVEMVKFAKPEESYQEELE 299
 RDEI+ G LPI + A S FRSEAGSAGRDTRGLIR HQF+KVE+VKF KPE+SY+ELE
 Sbjct: 241 RDEILSGDSLPI NYAAFSACFRSEAGSAGRDTRGLIRHQFKNVELVKFVKPEDSYEELE 300

20 Query: 300 KMTANAENILQKLNLPYRVITLCTGDMGFSAAKTYDLEVWIPAQNTYREISSCSNTEDFQ 359
 K+T AE +LQ L LPYRV+++CTGD+GF+AAK YD+EVWIP+Q+TYREISSCSN E FQ
 Sbjct: 301 KLINQAERVLQLEL PYRVMSMCTGDLGFTA AKKYDIEVWIPSQD TYREISSCSNFEAFQ 360

25 Query: 360 ARRAQIRYRDEVDGKVRLLHTLNGSGLAVGRTVAAILENYQEDGSVTIPEVLRPYMGNI 419
 ARRA IR+R E GK +HTLNGSGLAVGRTVAAILENYQ EDGSV IP+VLRPYMGN
 Sbjct: 361 ARRANIRFRREAKGKPEHVHTLNGSGLAVGRTVAAILENYQEDGSVVIPKVLRPYMGNR 420

25 Query: 420 DIIKP 424
 +++KP
 Sbjct: 421 EVMKP 425

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5225> which encodes the amino acid
 30 sequence <SEQ ID 5226>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2453(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 357/424 (84%), Positives = 386/424 (90%)

40 Query: 1 MLDLKRIRTDVVDVAKKLATRGVDQETLTTLKELDIKRRELLIKAEEAKAQRNVASAAIA 60
 MLDLKRIRTDV D VA KL RGV ++TLT LKELD KRR LL+++EE KA+RN+ASAAIA
 Sbjct: 1 MLDLKRIRTDVDTVA AKLKNRGVSEDTLTHLKELEKRRALLVQSEELKAERNIASAAIA 60

45 Query: 61 QAKRNKENADEQIAAMQTLASADIKADAELADVANDLQSMVTVLPNTPADDPVPLGADEDE 120
 QAKR KE+A +QIA MQ +SADIK ID +L +D + ++TVLPNTP D VP+GADE++
 Sbjct: 61 QAKRQKEDATQQIADMQKVSADIKTIDNQLVAIDQQVTDIITVLPNTPHDSVPVGADEED 120

50 Query: 121 NVEVRRWGTTPREFDFETKAHWDLGESLGILDWERGAKVTGSRFLFYKGLGARLERAIYSF 180
 NVE+RRWGTTPR+FD FE KAHWDLGE L ILDWERGAKVTG+RFLFYK LGARLERAIY+F
 Sbjct: 121 NVEIRRWGTTPRDFDFEVKAHWDLGEDLDILDWERGAKVTGARFLFYKNL GARLERALYNF 180

55 Query: 181 MLDEHAKEGYTEVIPPYVMVNHDSMFGTGQYPKFKEDTFELADSPFVLIPTAEVPLTNY 240
 MLDEH KEGY E+I PYMVNHDSMFGTGQYPKFKEDTFELAD+ FVLIPTAEVPLTNY
 Sbjct: 181 MLDEHIKEGYQEIIITPYMVNHDSMFGTGQYPKFKEDTFELADTNFVLIPTAEVPLTNY 240

60 Query: 241 DEIIDGKELPIYFTAMSPSFRSEAGSAGRDTRGLIRLHQFHKVEMVKFAKPEESYQEELEK 300
 EI+DGKELPIYFTAMSPSFRSEAGSAGRDTRGLIRLHQFHKVEMVKFAKPEESYQEELEK
 Sbjct: 241 GEILDGKELPIYFTAMSPSFRSEAGSAGRDTRGLIRLHQFHKVEMVKFAKPEESYQEELEK 300

65 Query: 301 MTANAENILQKLNLPYRVITLCTGDMGFSAAKTYDLEVWIPAQNTYREISSCSNTEDFQA 360
 MTANAENILQKL LPYRVI+LCTGDMGFSAAKTYDLEVWIPAQNTYREISSCSNTEDFQA
 Sbjct: 301 MTANAENILQKLGLPYRVISLCTGDMGFSAAKTYDLEVWIPAQNTYREISSCSNTEDFQA 360

-1896-

Query: 361 RRAQIRYRDEVDGKVRLLHTLNGSGLAVGRTVAAILENYQNEGDSVTIPEVLRPYMGNIID 420
 RRAQIRYRDE DGKV+LLHTLNGSGLAVGRTVAAILENYQNEGDSVTIPEVLRPYMG
 Sbjct: 361 RRAQIRYRDEADGKVKLLHTLNGSGLAVGRTVAAILENYQNEGDSVTIPEVLRPYMGGET 420

5 Query: 421: IIKP 424
 +I P
 Sbjct: 421 VISP 424

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 1684

A DNA sequence (GBSx1788) was identified in *S.agalactiae* <SEQ ID 5227> which encodes the amino
 acid sequence <SEQ ID 5228>. Analysis of this protein sequence reveals the following:

Possible site: 36
 15 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.36	Transmembrane	313 - 329 (306 - 332)
INTEGRAL	Likelihood = -9.24	Transmembrane	159 - 175 (155 - 179)
INTEGRAL	Likelihood = -4.19	Transmembrane	20 - 36 (16 - 37)
INTEGRAL	Likelihood = -3.29	Transmembrane	271 - 287 (271 - 287)
20 INTEGRAL	Likelihood = -2.97	Transmembrane	210 - 226 (209 - 227)
INTEGRAL	Likelihood = -2.87	Transmembrane	242 - 258 (241 - 258)
INTEGRAL	Likelihood = -2.13	Transmembrane	52 - 68 (50 - 68)

----- Final Results -----
 25 bacterial membrane --- Certainty=0.5543(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9559> which encodes amino acid sequence <SEQ ID 9560>
 30 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA07406 GB:AJ006986 transmembrane protein [Streptococcus pneumoniae]
 Identities = 72/330 (21%), Positives = 143/330 (42%), Gaps = 32/330 (9%)

35 Query: 14 RHYGLDLLRIISMFMIVITHVLGKGLRSSVEGHADSYFIVTWIIQVLVYGAVNCYALIS 73
 R+ LDLL++++ +V+ H GG + + + +Y + ++ VN Y L+
 Sbjct: 5 RNINLDDLKVLACVGVLLHTT-MGGFKETGAWNFLTLYLYLGTYSIPLFFMVNGYLL- 62

40 Query: 74 GYVGINSRYRSKLLSIWQVFFYFTTITLFAITGHE-----VTLLNWRDAFFPIVSG 127
 G I Y K+ + V +TF I LF E + L + FF
 Sbjct: 63 GKREITYSYILQIKWLLITVSSWTF-IVWLFKRDTENLIKIIIGSLIQGYFF----- 116

45 Query: 128 QYWYITAYFGLLVFMPVINNGLNALTDKQLKQLVLLMFI--IFSILPAVLNNRVPEFSLS 185
 Q+W+ A + + +P++ LN+ L L LLM I IF + +L + + +
 Sbjct: 117 QFWFFGALILIYLCLPILRQFLNS-KRSYLYSLSLMTIGLIFELSNILLQMPIQTYVIQ 175

50 Query: 186 KGFEMTWLLILYIIGAYLKRIDL---NIFKTSYLLIYLLSLVATYAMKFSVGDIV--- 238
 TW Y++G Y+ + + + FK ++ LL L++ + F I+
 Sbjct: 176 TFRLLTW-FFYLLGGYIAQFTIEEIESRFKNWMKIVSILLLLISPIILFFIAKTIYHNL 234

55 Query: 239 ---YWYVSPTLTLGAVSLFILFARASIKPSGFLKKIIVVLAPSTLGVYLCHLHPLIVKYF 295
 Y+Y + + + + +F+ ++ + ++ IV L+ T+GV++ +H I+K +
 Sbjct: 235 FAEYFYDTLFFVKVSTLGIFLTILMLTLNEN--RRESIVSLSNQTMGVFI--IHTYIMKWV 290

Query: 296 VRDFAETFFVYESIYLPFLILGAGILIYLL 325
 + FV + F + + I++ +L
 Sbjct: 291 EKVLGFNFVGAYLLFALFTLSVSFIIVGML 320

No corresponding DNA sequence was identified in *S.pyogenes*.

-1897-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1685

A DNA sequence (GBSx1789) was identified in *S.agalactiae* <SEQ ID 5229> which encodes the amino acid sequence <SEQ ID 5230>. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2752(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9561> which encodes amino acid sequence <SEQ ID 9562> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD46488 GB:AF130465 unknown [Streptococcus salivarius]
Identities = 88/112 (78%), Positives = 96/112 (85%)

20 Query: 1 MAQSLNKTVEFQTTGVSYLGMGNKVGKFLVGDQALEFYNDKNVNDYI QIPWTSINQIGAN 60
 MAQSLNKTVE TTGVSY+ +G KVGKFL+GD ALEFY D NV YIQIPWTSI QIGAN
 Sbjct: 1 MAQSLNKTVELHTTGVSYMAIGGKVGKFLIGDVALEFYPDVNVEQYI QIPWTSITQIGAN 60

25 Query: 61 VSRKKISRHFVFTDQGKFLFASKDSGTILKHARRHIGDDKVVKLPTLIQTI 112
 VS K+ISRHFV TD+ KFLFASKDSG ILK AR H+G++KVVKLPTLIQTI
 Sbjct: 61 VSGKRISRHFVLTDKSKFLFASKDSGKILKIAREHLGNEKVVKLPTLIQTI 112

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5231> which encodes the amino acid sequence <SEQ ID 5232>. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.3301(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 87/116 (75%), Positives = 101/116 (87%)

40 Query: 1 MAQSLNKTVEFQTTGVSYLGMGNKVGKFLVGDQALEFYNDKNVNDYI QIPWTSINQIGAN 60
 MAQSLN +VE++T VSYLGMG KVG L+GD+ALEFYNDKNVNDYI QIPWT+IN IGAN
 Sbjct: 1 MAQSLNTSVEYKTKAVSYLGMGGKVGHILGDKALEFYNDKNVNDYI QIPWTAINHIGAN 60

45 Query: 61 VSRKKISRHFVFTDQGKFLFASKDSGTILKHARRHIGDDKVVKLPTLIQTI LKIF 116
 VSRKK+SRHFE+FTDQGKFLFAS DSG ILK R+HIG++KV+ LPTL+QT + F
 Sbjct: 61 VSRKKVSRHFEIFTDQGKFLFASGDSGKILKITRQHIGNEKVITLPTLMQT FINKF 116

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1898-

Example 1686

A DNA sequence (GBSx1790) was identified in *S.galactiae* <SEQ ID 5233> which encodes the amino acid sequence <SEQ ID 5234>. This protein is predicted to be mannose-specific phosphotransferase system component IID (manZ). Analysis of this protein sequence reveals the following:

```

5   Possible site: 39
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -8.92    Transmembrane  281 - 297 ( 279 - 302)
      INTEGRAL    Likelihood = -4.88    Transmembrane  187 - 203 ( 185 - 205)
      INTEGRAL    Likelihood = -4.35    Transmembrane  260 - 276 ( 257 - 277)
10   INTEGRAL    Likelihood = -1.01    Transmembrane  129 - 145 ( 129 - 145)

      ----- Final Results -----
                        bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
                        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAD46487 GB:AF130465 mannose-specific phosphotransferase system
      component IID [Streptococcus salivarius]
20   Identities = 247/303 (81%), Positives = 276/303 (90%)

Query: 1   MTEQIKLSKSDRQKVWWSQFLOGSWNYERMQNMGWAYALIPALKKLYTTKEDRAAALER 60
      M E+I+LS++DR+KVWWSQFLOGSWNYERMQ+GWAY+LIPA+KKLYT KED+AAAL+R
Sbjct: 1   MAEKIQLSQADRKKVWWSQFLOGSWNYERMQNLGWAYSLIPAICKLYTNKEDQAAALKR 60

25   Query: 61 HMEFFNTHPYVAAPIIGVTLALEEEKASGTPVEDKAIQGVKIGMMGPLAGIGDPVFWFTV 120
      H+EFFNTHPYVAAPI+GVTLALEEEKA+GT +ED AIQGVKIGMMGPLAGIGDPVFWFTV
Sbjct: 61 HLEFFNTHPYVAAPIMGVTLALEEEKANGTDIEDAAIQGVKIGMMGPLAGIGDPVFWFTV 120

30   Query: 121 RPILGALGASLASAGNILGPPIFFVGWNLIRMSFLWYTQELGYKSGKEITKDMSSGGILQD 180
      RPILGALGASLA AGNI GP+IFF+GWNLIRM+FLWYTQELGYK+G EITKDMSSGGIL+D
Sbjct: 121 RPILGALGASLAQAGNIAGPLIFFIGWNLIRMAFLWYTQELGYKAGSEITKDMSSGGILKD 180

35   Query: 181 ITKGASILGMFILAVLVKRWVAINFTVDLPKKTLSSEGAYINFPKDHVSGQQLHDILGQVQ 240
      ITKGASILGMFILAVLV+RWV+I FTV+LP K LS+GAYI +PK +VSG QL ILGQV
Sbjct: 181 ITKGASILGMFILAVLVERWVSIVFTVNLPGKVLKSGAYIEWPKGNVSGDQLKTILGQVN 240

Query: 241 SGLSLDKMQPQTLQGQLDLSLIPGLAGLLLTFFCMWLLKKKVSPITIIIGLFIVGILARLA 300
      LS DK+Q TLQ QLDLSLIPGL GLLLTFF CMWLLKKKVSPITIIIGLF+VGI+A
40   Sbjct: 241 DKLSFDKIQVDTLQQLDLSLIPGLMGLLLTFACMWLLKKKVSPITIIIGLFVVGIVASFF 300

Query: 301 GVM 303
      G+M
Sbjct: 301 GIM 303
45

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5235> which encodes the amino acid sequence <SEQ ID 5236>. Analysis of this protein sequence reveals the following:

```

      Possible site: 55
   >>> Seems to have no N-terminal signal sequence
80   INTEGRAL    Likelihood = -8.39    Transmembrane  284 - 300 ( 279 - 302)
      INTEGRAL    Likelihood = -4.88    Transmembrane  261 - 277 ( 257 - 278)
      INTEGRAL    Likelihood = -4.51    Transmembrane  181 - 197 ( 180 - 198)

      ----- Final Results -----
55   bacterial membrane --- Certainty=0.4354(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

60   >GP:AAD46487 GB:AF130465 mannose-specific phosphotransferase system
      component IID [Streptococcus salivarius]

```

-1899-

Identities = 239/303 (78%), Positives = 268/303 (87%)

Query: 1 MTEQIKLTKSDRQVRVWRSQFLQGSWNYERMQNMGWAYALIPALKKLYTSPEDRAAALER 60
 M E+I+L+++DR++VWWSRQFLQGSWNYERMQNGWAY+LIPA+KKLYT+ ED+AAAL+R
 5 Sbjct: 1 MAEKIQLSQADRKKVWWSRQFLQGSWNYERMQNLGWAYSLIPAICKLYTNKEDQAAALKR 60

Query: 61 HMEFFNTHPYVAAPIIGVTLALEEERANGTPIDDKAIQGVKIGMMGPLAGIGDPVFWFTI 120
 H+EFFNTHPYVAAPI+GVTLALEEE+ANGT I+D AIQGVKIGMMGPLAGIGDPVFWFT+
 10 Sbjct: 61 HLEFFNTHPYVAAPIMGVTLALEEEKANGTDIEDAAIQGVKIGMMGPLAGIGDPVFWFTV 120

Query: 121 RPILGALGASLASTGNIVGPLLFFFGWNLIRMAFLWYTQEFQYKAGSEITKDMSSGGILQD 180
 RPILGALGASLA GNI GPL+FF GWNLIRMAFLWYTQE GYKAGSEITKDMSSGGIL+D
 Sbjct: 121 RPILGALGASLAQAGNIAGPLIFFIGWNLIRMAFLWYTQELGYKAGSEITKDMSSGGILKD 180

Query: 181 ITKGASILGMFILAVLVQRVWSINFTIDLPGKQLSDGAYVVPDGA VKAELKILANAI 240
 ITKGASILGMFILAVLV+RWVSI FT++LPGK LS GAY+ +P G V G +LKITIL
 15 Sbjct: 181 ITKGASILGMFILAVLVERVWSIVFTVNLPGKVLKSGAYIEWPKGNVSGDQLKITILGQVN 240

Query: 241 GGMSLDKVAQTLQGLDLSLIPGLAGLLLTFLCMWLLKKKVSPITIIIGLFAFGILAHLA 300
 +S DK+Q TLQ QLDLSLIPGL GLLLTFL CMWLLKKKVSPI IIIGLF GI+A
 20 Sbjct: 241 DKLSFDKIQVDTLQKQLDLSLIPGLMGLLLTFLACMWLLKKKVSPITIIIGLFFVVGIVASFF 300

Query: 301 GIM 303
 GIM
 25 Sbjct: 301 GIM 303

An alignment of the GAS and GBS proteins is shown below.

Identities = 255/303 (84%), Positives = 277/303 (91%)

Query: 1 MTEQIKLSKSDRQKVVWRSQFLQGSWNYERMQNMGWAYALIPALKKLYTTKEDRAAALER 60
 MTEQIKL+KSDRQ+VWWSRQFLQGSWNYERMQNMGWAYALIPALKKLYT+ EDRAAALER
 30 Sbjct: 1 MTEQIKLTKSDRQVRVWRSQFLQGSWNYERMQNMGWAYALIPALKKLYTSPEDRAAALER 60

Query: 61 HMEFFNTHPYVAAPIIGVTLALEEEKASGTPVEDKAIQGVKIGMMGPLAGIGDPVFWFTV 120
 HMEFFNTHPYVAAPIIGVTLALEEE+A+GTP++DKAIQGVKIGMMGPLAGIGDPVFWFT+
 35 Sbjct: 61 HMEFFNTHPYVAAPIIGVTLALEEERANGTPIDDKAIQGVKIGMMGPLAGIGDPVFWFTI 120

Query: 121 RPILGALGASLASAGNIGPIIFFVGNLIRMSFLWYTQELGYKSGKEITKDMSSGGILQD 180
 RPILGALGASLAS GNI+GP++FF GWNLIRM+FLWYTQE GYK+G EITKDMSSGGILQD
 40 Sbjct: 121 RPILGALGASLASTGNIVGPLLFFFGWNLIRMAFLWYTQEFQYKAGSEITKDMSSGGILQD 180

Query: 181 ITKGASILGMFILAVLVKRWVAINFTVDLPKKTLSSEGAYINFPKDHVSGQQLHDILGQVQ 240
 ITKGASILGMFILAVLV+RWV+INFT+DLP K LS+GAY+ FP V G +L IL
 45 Sbjct: 181 ITKGASILGMFILAVLVQRVWSINFTIDLPGKQLSDGAYVVPDGA VKAELKILANAI 240

Query: 241 SGLSLDKMQPQTLQGLDLSLIPGLAGLLLTFFCMWLLKKKVSPITIIIGLFIVGILARLA 300
 G+SLDK+Q QTLQGLDLSLIPGLAGLLLTFL CMWLLKKKVSPI IIIGLF GILA LA
 50 Sbjct: 241 GGMSLDKVAQTLQGLDLSLIPGLAGLLLTFLCMWLLKKKVSPITIIIGLFAFGILAHLA 300

Query: 301 GVM 303
 G+M
 55 Sbjct: 301 GIM 303

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 55 vaccines or diagnostics.

Example 1687

A DNA sequence (GBSx1791) was identified in *S.agalactiae* <SEQ ID 5237> which encodes the amino
 acid sequence <SEQ ID 5238>. Analysis of this protein sequence reveals the following:

Possible site: 22
 60 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1900-

bacterial cytoplasm --- Certainty=0.2580(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1688

10 A DNA sequence (GBSx1792) was identified in *S.agalactiae* <SEQ ID 5239> which encodes the amino acid sequence <SEQ ID 5240>. This protein is predicted to be mannose-specific phosphotransferase system component IIC (manY). Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have a cleavable N-term signal seq.

15 INTEGRAL Likelihood = -5.95 Transmembrane 142 - 158 (137 - 165)
 INTEGRAL Likelihood = -2.60 Transmembrane 65 - 81 (61 - 81)
 INTEGRAL Likelihood = -1.97 Transmembrane 103 - 119 (103 - 122)

----- Final Results -----

20 bacterial membrane --- Certainty=0.3378(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 A related GBS nucleic acid sequence <SEQ ID 9301> which encodes amino acid sequence <SEQ ID 9302> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD46486 GB:AF130465 mannose-specific phosphotransferase system
 component IIC [Streptococcus salivarius]

Identities = 134/186 (72%), Positives = 154/186 (82%), Gaps = 1/186 (0%)

30 Query: 1 MVKSGDFTQKGINFAFSTAVPLAIAAGLFLTMIVRTIISTALVHAGDKAASEGNFAAIERFH 60
 +VK G+FT +GI A +TA+PLA+AGLFLTM+VRT S ALVHA DKAA GN A +ER H
 Sbjct: 86 LVKGGNFTTEGIGVATATAIPLAVAGLFLTMLVRTASVALVHAADKAAESGNIAGVERAH 145

35 Query: 61 FIALLLQGLRIAFFPAALLLAIPSSSVQSILEAMPDWLNGGMQVGGAMVVAVGYAMVINMM 120
 ++ALLLQGLRIA PAALLLAIP+ SVQ L MP WLN GM VGG MVA VGYAMVINMM
 Sbjct: 146 YLALLLQGLRIAVPAALLLAIPAESVQHGLMPSWLNHGMVGGGMVVAVGYAMVINMM 205

40 Query: 121 ATREVWPPFFALGFALAAALNQLTLIAMGTIGVAIALIYISLSKMGGSK-GTSNAGSNDPIG 179
 ATREVWPPFFA+GFA AA++QLTLIA+G IGVAIA IY++LSK GG G +++GS DFIG
 Sbjct: 206 ATREVWPPFAIGFAFAAISQLTLIALGAIGVAIAFIYLNLSKQGGGNGGGTSSGSGDPIG 265

Query: 180 DILEDY 185
 DILEDY

45 Sbjct: 266 DILEDY 271

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5241> which encodes the amino acid sequence <SEQ ID 5242>. Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have an uncleavable N-term signal seq

50 INTEGRAL Likelihood = -11.30 Transmembrane 4 - 20 (1 - 28)
 INTEGRAL Likelihood = -7.64 Transmembrane 226 - 242 (212 - 247)
 INTEGRAL Likelihood = -4.14 Transmembrane 102 - 118 (101 - 123)
 INTEGRAL Likelihood = -3.77 Transmembrane 71 - 87 (69 - 87)
 55 INTEGRAL Likelihood = -3.40 Transmembrane 150 - 166 (146 - 167)

-1901-

INTEGRAL Likelihood = -2.13 Transmembrane 186 - 202 (186 - 202)
 INTEGRAL Likelihood = -0.37 Transmembrane 37 - 53 (37 - 53)

----- Final Results -----

5 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AAD46486 GB:AF130465 mannose-specific phosphotransferase system
 component IIC [Streptococcus salivarius]
 Identities = 211/271 (77%), Positives = 237/271 (86%), Gaps = 2/271 (0%)

15 Query: 1 MSDISIISAILVVIIAFFAGLEGILDQFQMHQPLVACTLIGLVTGHLEAGVILGGTLQML 60
 MSD+SIISAILVV++AF AGLEGILDQFQ HQPLVACTLIG TG+L AG++LGG+LQM+
 Sbjct: 1 MSMSIISAILVVVVAFLAGLEGILDQFQFHQPLVACTLIGAATGNLTAGIMLGGSLQMI 60

Query: 61 ALGWANIGAAVAPDAALASVAAAIIMVKS GDFQKGITFAYSTAIP LAVAGLFLTMIVRT 120
 AL WANIGAAVAPDAALASVAAAI+VK G+FT +GI A +TAIP LAVAGLFLTM+VRT
 20 Sbjct: 61 ALAWANIGAAVAPDAALASVAAAIILVKG NFTTEGIGVATATAIP LAVAGLFLTMLVRT 120

Query: 121 LSTALVHAGDKAAAE GNFAGIERFHFIALLLQGLRIAVPAALLVAVPTS AVQSVLNAMPN 180
 S ALVHA DKAA GN AG+ER H++ALLLQGLRIAVPAALL+A+P +VQ L MP+
 25 Sbjct: 121 ASVALVHAADKAAESGNIAGVERAHY LALLLQGLRIAVPAALLLAIPAESVQHAGLMPN 180

Query: 181 WLN EGMQIGGAMVVAVGYAMVINMMATREVWPF FALGFALAAISQLTLIAMGVIGVAIAF 240
 WLN GM +GG MVVAVGYAMVINMMATREVWPF FA+GFA AAISQLTLIA+G IGVAIAF
 30 Sbjct: 181 WLN HGMVVGGMVVAVGYAMVINMMATREVWPF FAIGFAFAAISQLTLIALGAIGVAIAF 240

Query: 241 IYLNLSKKG G--NGGNAAGSADPIGDILEDY 269
 IYLNLSK+GG GG ++GS DPIGDILEDY
 35 Sbjct: 241 IYLNLSKQGGGNGG TSSGSGDPIGDILEDY 271

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 155/185 (83%), Positives = 173/185 (92%), Gaps = 1/185 (0%)

Query: 1 MVKSGDFTQKGINFAFSTAVPLAAGLFLTMIVRTISTALVHAGDKAAEGNFAA IERFH 60
 MVKSGDFTQKGI FA+STA+PLA+AGLFLTMIVRT+STALVHAGDKAA+EGNFA IERFH
 40 Sbjct: 86 MVKSGDFTQKGITFAYSTAIP LAVAGLFLTMIVRTLSTALVHAGDKAAAE GNFAGIERFH 145

Query: 61 FIALLLQGLRIAFPAALLLAIPSSSVQSILEAMPDNLNGGMQVGGAMVVAVGYAMVINMM 120
 FIALLLQGLRIA PAALL+A+P+S+VQS+L AMP+WLN GMQ+GGAMVVAVGYAMVINMM
 45 Sbjct: 146 FIALLLQGLRIAVPAALLVAVPTS AVQSVLNAMPNWLNEGMQIGGAMVVAVGYAMVINMM 205

Query: 121 ATREVWPF FALGFALAAINQLTLIAMGTIGVAIALIYISLSKMGGSKGTSNAGSNDPIGD 180
 ATREVWPF FALGFALAA++QLTLIAMG IGVAIA IY++LSK GG+ G + AGS DPIGD
 50 Sbjct: 206 ATREVWPF FALGFALAAISQLTLIAMGVIGVAIAFIYLNLSKKGNGGNA-AGSADPIGD 264

Query: 181 ILEDY 185
 ILEDY
 55 Sbjct: 265 ILEDY 269

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

55 Example 1689

A DNA sequence (GBSx1793) was identified in *S.agalactiae* <SEQ ID 5243> which encodes the amino acid sequence <SEQ ID 5244>. Analysis of this protein sequence reveals the following:

possible site: 37
 >>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----

-1902-

bacterial cytoplasm --- Certainty=0.3171(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1690

10 A DNA sequence (GBSx1794) was identified in *S.agalactiae* <SEQ ID 5245> which encodes the amino acid sequence <SEQ ID 5246>. This protein is predicted to be pseudouridine synthase (rluC). Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2717(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06566 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 124/281 (44%), Positives = 171/281 (60%), Gaps = 8/281 (2%)

Query: 16 LLKSHDVSRLAKIKYRGKKIFVNGEEQNAIFLLEIGDVVTIDIPDE-PSHETL-EPVP 73
 L + VS+ LA IK++GG I +NGEE + + D VT+++P E PS + EPVP
 Sbjct: 24 LREGKHVSKRSLAAIKFKGGTILLNGEEVTVRETVDVNDQVTLLELPHEYFSPSMIAEPVP 83

Query: 74 HDLDIIYEDDHFLILNKPFASFIPSSIH-SNTIANFIKHYVSNNYANQQVHIVTRLDR 132
 D+IYE+DH+L++NKP G +IPS H T+AN + +Y+ A H V RLD+
 Sbjct: 84 --FDVIYENDHYLVVNKPAGVPTIPSRDHPQGTLANGLLNYFQRQKMA-ATFHAVNRLDK 140

Query: 133 DTSGMLLFAKHGYAHARLDKQLQAKAIEKRYVALVSGSGDLADSGDIAPARDVDSIIT 192
 DTSGL++ AKH AH +L KQ + I++ Y A+V G + + G I APIAR +S+IT
 Sbjct: 141 DTSGLLIVAKHQLAHDQLSKQQRQGNIKRTYMAIVQGEIEBQQE-GTITAPIARKEESLIT 199

Query: 193 RRVHESGKYAHTSYQVARYGDVRLVDIKLHTGRTHQIRVHFAHIGFPLLGDDLYGGRMD 252
 R V E G+ A T ++V+ R +V ++L TGRTHQIRVHF+++G+PL GDDLYGG
 Sbjct: 200 REVREDGQLAITHFKVIDRLNQGTIVQVQLETGRTHQIRVHFSYLGYPFLFGDDLYGGERK 259

Query: 253 LGINRQALHCHSLSFYDPFMGKINKQTLDLTDDFDSVIMEL 293
 GI RQALH L+ + PF T L D +I L
 Sbjct: 260 -GIERQALHSTELTIHCPFTEVEQTFTTEGLPPDMKELIRHL 299

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5247> which encodes the amino acid sequence <SEQ ID 5248>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2786(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 An alignment of the GAS and GBS proteins is shown below.

Identities = 223/294 (75%), Positives = 251/294 (84%), Gaps = 1/294 (0%)

-1903-

Query: 1 MKFEYVAKERCKVKITLLKSHDVSRLAKIKYRGKIFVNGEEQNAIFLLEIGDVVTIDI 60
 M+FE+VA +R KVKITLLKS+DVS+GLLAKIKY+GG I VNG EQNAI+LL++GDVVTIDI
 Sbjct: 1 MRFEFVADKRIKVKITLLKSYDVSGLLAKIKYKGGNILLVNGIEQNAIYLLQVGDVVTIDI 60

5 Query: 61 PDEPSHETLEPVPHDLDIYEDDHFLILNKPFGFASIPSSIHSNTIANFIKHYVSNNYA 120
 P+E E LE +P DLDI++EDDHFL++NKP GFASIPS+IHSNTIANFIK YV N+Y
 Sbjct: 61 PNEEPFEKLEAIPFDLDIVHEDDHFLVINKPIGFASIPSAIHSNTIANFIKAYVDNHYL 120

10 Query: 121 NQQVHIVTRLDRDTSGMLFAKHGYAHARLDKQLQAKAIEKRYALVSGSGDLADSGDII 180
 +QQVHIVTRLDRDTSGMLFAKHGYAHARLDKQLQ ++IEKRY+ALVSG+G L D GDII
 Sbjct: 121 DQQVHIVTRLDRDTSGMLFAKHGYAHARLDKQLQTRSIEKRYFALVSGNGMLPDEGDII 180

15 Query: 181 APIARDVDSIITRRVHESGKYAHTSYQVVARYGD-VRLVDIKLHTGRTHQIRVHFAHIGF 239
 API R DSIITR V GKYA TSY+VVARY + V LVDIKLHTGRTHQIRVHFAHIGF
 Sbjct: 181 APIGRSKDSIITRAVDPMGKYAKTSYKVARYSENVHLVDIKLHTGRTHQIRVHFAHIGF 240

20 Query: 240 PLLGDDLYGGRMDLGINRQALHCHSLSFYDPFMGKINKQTLDTDDFDSVIMEL 293
 PLLGDDLYGGR+DLGI RQALHCH L+F DPF + LTDDFDSVI+ L
 Sbjct: 241 PLLGDDLYGGRDLGITRQALHCHYLNFKDPFTESDCSYAIHLTDDFDSVIIGL 294

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1691

25 A DNA sequence (GBSx1795) was identified in *S. agalactiae* <SEQ ID 5249> which encodes the amino acid sequence <SEQ ID 5250>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1521(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 A related GBS nucleic acid sequence <SEQ ID 9845> which encodes amino acid sequence <SEQ ID 9846> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13018 GB:Z99110 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 120/267 (44%), Positives = 174/267 (64%), Gaps = 3/267 (1%)

40 Query: 13 RVAIIANGKYQSKRVASKLFAAFKHDPDFYLSKKOPDIVISIGDGMLLSAFHYEKQLD 72
 + A+ + G S + SK+ A+ D D L + +P+IVIS+GGDG LL AFH Y +LD
 Sbjct: 2 KFAVSSKGDQVSDTLKSKI-QAYLLDFDMELDENEPFIVISVGGDGTLLYAFHRYSDRLD 60

45 Query: 73 KVRFGVHTGHLGFYTDYRDFEVDTLINNLKNDKGEQISYPILKVTITL-EDGRVIRARA 131
 K FVGVTGHLGFY D+ E++ L+ + + YP+L+V +T E+ R R A
 Sbjct: 61 KTAFGVHTGHLGFYADWVPHEIEKLVLAIKTPYHTVEYPLLEVIVTYHENEREERYLA 120

50 Query: 132 LNESTIKRIEKTMDVADVQVVFERFRGDGILVSTPTGSTAYNKS LGGAVLHPTIEALQ 191
 LNE TIK IE ++VADV I +FE FRGDG+ +STP+GSTAYNK+LGGA++HP+I A+Q
 Sbjct: 121 LNECTIKSIEGSLVADVEIKGQLFETFRGDGLCLSTPSGSTAYNKALGGAIHPHRAIQ 180

55 Query: 192 LTEISSLNNRVYRTLGSSVPIPKDAIEIVPKRVGVYTTISIDNKTVHYKNVTKIEYSIDE 251
 L E++S+NNRV+RT+GS +++P I P+ + ++ID+ T+ +K+V I +
 Sbjct: 181 LAEMASINNRVFRTVGSPLLLPSHDCMIKPRNEVDFOVTIDHLTLHLKDVKSIRCQVAS 240

60 Query: 252 KSINFVSTPSHTSFWERVNDAFIGEPE 278
 + + F FW+RV D+FIG+ E
 Sbjct: 241 EKVRFRFRPF-PFWKRVQDSFIGKE 266

-1904-

A related sequence was also identified in GAS <SEQ ID 9137> which encodes the amino acid sequence <SEQ ID 9138>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2190(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 155-157

An alignment of the GAS and GBS proteins is shown below.

Identities = 232/276 (84%), Positives = 257/276 (93%)

Query: 1 MMTQMNFTDRATRVAILIANGKYQSKRVASKLFAAFKHDPDFYLSKKDPDIVISIGGDGML 60

+MTQMN+T + RVAIIANGKYQSKRVASKLF+ FK DPDFYLSKK+PDIVISIGGDGML

Sbjct: 1 VMTQMNYTGKVKRVAILIANGKYQSKRVASKLFSVFKDDPDFYLSKKNPDIVISIGGDGML 60

Query: 61 LSAFHMYEKQLDKVRFVGVHTGHLGFYTDYRDFEVDTLINNLKNDKGEQISYPILKVTIT 120

LSAFHMYEK+LDKVRVFGV+HTGHLGFYTDYRDFEVD LI+NL+ DKGEQISYPILKV IT

Sbjct: 61 LSAFHMYEKELDKVRFVGIHTGHLGFYTDYRDFEVDKLIDNLRKDKGEQISYPILKVAIT 120

Query: 121 LEDGRVIRARALNESTIKRIEKTADVADVQVFERFRGDGILVSTPTGSTAYNKSLLGG 180

L+DGRV++ARALNE+T+KRIEKTADVADV+IN V FE FRGDGI VSTPTGSTAYNKSLLGG

Sbjct: 121 LDDGRVVKARALNEATVKRIEKTADVADV+INHVKFESFRGDGISVSTPTGSTAYNKSLLGG 180

Query: 181 AVLHPTIEALQLTEISSLNRRVYRTLGSSVVIIPKDAIEIVPKRVGVYTISIDNKTVHYK 240

AVLHPTIEALQLTEISSLNRRV+RTLGSS+IIPKDI IE+VPRK+G+YTISIDNKT K

Sbjct: 181 AVLHPTIEALQLTEISSLNRRVFTLGSSVVIIPKDKIELVPRKLGIVYTISIDNKTQYQLK 240

Query: 241 NVTKEYSIDEKSINFVSTPSHTSFWERVNDFAFIGE 276

NVTK+EY ID++ I+FVS+PSHTSFWERV DAFIGE

Sbjct: 241 NVTKVEYFIDDEKIHVSSPSHTSFWERVKDAFIGE 276

A related GBS gene <SEQ ID 8879> and protein <SEQ ID 8880> were also identified. Analysis of this protein sequence reveals an RGD motif at residues 159-161.

The protein has homology with the following sequences in the databases:

45.0/65.6% over 264aa

Bacillus subtilis

EGAD|107338| hypothetical protein Insert characterized OMNI|NT01BS1363 BC541A protein-related Insert characterized

SP|O31612|YJBN_BACSU HYPOTHETICAL 30.0 KDA PROTEIN IN MECA-TENA INTERGENIC REGION. Insert characterized

GP|2633515|emb|CAB13018.1||Z99110 similar to hypothetical proteins Insert characterized

PIR|F69844|F69844 conserved hypothetical protein yjbN - Insert characterized

ORF02026(337 - 1134 of 1437)

EGAD|107338|BS1162(2 - 266 of 266) hypothetical protein {Bacillus subtilis} OMNI|NT01BS1363 BC541A protein-related SP|O31612|YJBN_BACSU HYPOTHETICAL 30.0 KDA PROTEIN IN MECA-TENA INTERGENIC REGION. GP|2633515|emb|CAB13018.1||Z99110 similar to hypothetical proteins {Bacillus subtilis} PIR|F69844|F69844 conserved hypothetical protein yjbN - Bacillus subtilis

%Match = 22.8

%Identity = 44.9 %Similarity = 65.5

Matches = 120 Mismatches = 89 Conservative Sub.s = 55

87 117 147 177 207 237 267 297

RKF*QKYKSELWL*IFGQPSNIH*ITSIRGTSLKLNKDWKQKQSL*NWMKKCVRFKIFVKHSFYLLI*IEN*AMV*E

327 357 387 417 447 477 507 537

[illegible]

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence
```

```

bacterial cytoplasm --- Certainty=0.2190(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

Score = 481 bits (1224), Expect = e-138
Identities = 233/276 (84%), Positives = 257/276 (92%)

```

Query: 1 VMTQMNYTGKVKRVAI IANGKYQSKRVASKLFSVFKDDPDFYLSKKNPDIVISIGGDGML 60
VMTQMN+T + RVAI IANGKYQSKRVASKLF+ FK DPDFYLSKK+PDIVISIGGDGML
Sbjct: 1 VMTQMNFTRATRVAI IANGKYQSKRVASKLFAAFKHDPDFYLSKKDPDIVISIGGDGML 60

Query: 61 LSAFHMYEKELDKVRVFGIHTGHLGFTYDYRDFEVDKLI DNLRKDKEQISYPILKVAIT 120
LSAFHMYEK+LDKVRVFG+HTGHLGFTYDYRDFEVD LI+NL+ DKGEQISYPILKV IT
Sbjct: 61 LSAFHMYEKQLDKVRVFGVHTGHLGFTYDYRDFEVDTLINNLKNDKGEQISYPILKVIT 120

Query: 121 LDDGRVVKARALNEATVKRIEKT MVADVIINHVKFESFRGDGISVSTPTGSTAYNKS LGG 180
L+DGRV++ARALNE+T+KRIEKT MVADV+IN V FE FRGDGI VSTPTGSTAYNKS LGG
Sbjct: 121 LEDGRVIRARALNESTIKRIEKT MVADV VINQVFERFRGDGILVSTPTGSTAYNKS LGG 180

Query: 181 AVLHPTIEALQLTEISSLNRRVFTLGS SIIIPKKDKIELVPKRLGIYTTISIDNKT VQLK 240
AVLHPTIEALQLTEISSLNRRV+RTLGS S+IIPKKD IE+VPKR+G+YTTISIDNKT K
Sbjct: 181 AVLHPTIEALQLTEISSLNRRVRYRTLGS SVIIPKKDAIEIVPKRVGYTTISIDNKT VHYK 240

Query: 241 NVTKVEYFIDDEKIH FVSSPSHTSFWERVKDAFIGE 276
NVTK+EY ID++ I+FVS+PSHTSFWERV DAFIGE
Sbjct: 241 NVTKIEYSIDEKSINFVSTPSHTSFWERVNDAFIGE 276

```

GBS308-GST was purified as shown in Figure 226, lane 8.

-1906-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1692

A DNA sequence (GBSx1796) was identified in *S.agalactiae* <SEQ ID 5253> which encodes the amino acid sequence <SEQ ID 5254>. This protein is predicted to be permease. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3653(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06568 GB:AP001516 GTP pyrophosphokinase [Bacillus halodurans]
Identities = 115/208 (55%), Positives = 159/208 (76%), Gaps = 3/208 (1%)

Query: 4 DWETFLDPYIQTVGELKIKLRGIRKQFRKQNRHSPIEFVTGRVKSVEISIQEKMVLRGISE 63
+W+ FL PY Q V ELK+KL+GIR+Q++K ++H+PIEFVTGRVK + SI +K + + I
Sbjct: 3 NWDVFLTPYKQAVEELKVKLKGIREQYQKSSKHTPIEFVTGRVKPISSILDKAIRKNIPL 62

Query: 64 ENLAQDLQDIAGLRIMVQFVDDVDEVLLLRKRDHMTVVQERDYITHMKSSGYRSYHVVV 123
+ L + +QD+AGLRI+ QFV+D++ V+ L+R R D +V+ERDY+ K SGYRSYH+V+
Sbjct: 63 DQLEEKMDLAGLRIVTQFVEDIETVVQLIRSRDFEIVEERDYVEQKDSGYRSYHLVL 122

Query: 124 EYPVDTIDGQKKVLAIEIQTTLAMNFWATIEHSLNYKYQGDFPEEIKQRLKTAIALEL 183
YPV TI+G+K++L E+QIRTLAMNFWATIEHSLNYKY G+ P IK RL++ A+ A L
Sbjct: 123 RYPVQTIEGEKRIIVELQIRTLAMNFWATIEHSLNYKYSGEIPLNIKTRLQRAEAAPRL 182

Query: 184 DEEMRKIREIREAQLLFDPLNRKLSDG 211
DEEM +IR+++REAQ + + RK G
Sbjct: 183 DEEMSQIRDEVREAQOI---ITRQEQG 207

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5255> which encodes the amino acid sequence <SEQ ID 5256>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4064(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 196/223 (87%), Positives = 213/223 (94%)

Query: 1 MSMDWETFLDPYIQTVGELKIKLRGIRKQFRKQNRHSPIEFVTGRVKSVEISIQEKMVLRG 60
M++DWE FLDPYIQTVGELKIKLRGIRKQ+RKQNR+SPIEFVTGRVKS+ESI+EKM+LRG
Sbjct: 1 MTLDWEEFLDPYIQTVGELKIKLRGIRKQYRKQNRYSPIEFVTGRVKSIESIKEKMILRG 60

Query: 61 ISEENLAQDLQDIAGLRIMVQFVDDVDEVLLLRKRDHMTVVQERDYITHMKSSGYRSYH 120
+ EEN+AQD+QDIAGLRIMVQFVDDV+EVLLLR+R DMT+V ERDYI +MKSSGYRSYH
Sbjct: 61 VIEENLAQDIQDIAGLRIMVQFVDDVEVLALLRQRQDMTIVYERDYIRNMKSSGYRSYH 120

Query: 121 VVVEYPVDTIDGQKKVLAIEIQTTLAMNFWATIEHSLNYKYQGDFPEEIKQRLKTAIA 180
VVVEYPVDTI+GQKKVLAIEIQTTLAMNFWATIEHSLNYKY GDFPEEIK+RLE TAKIA
Sbjct: 121 VVVEYPVDTIEGQKKVLAIEIQTTLAMNFWATIEHSLNYKYGDFPEEIKRLEVTAKIA 180

-1907-

Query: 181 LELDEEMRKIREDIRAQLLFDPLNRKLSGVDGNSDDTDEFYR 223
 LELDEEMRKIREDIRAQLLFDP+ R LSDGVDGNSDDTDE YR
 Sbjct: 181 LELDEEMRKIREDIRAQLLFDPVTRNLSGVDGNSDDTDELYR 223

5

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1693

A DNA sequence (GBSx1797) was identified in *S.agalactiae* <SEQ ID 5257> which encodes the amino acid sequence <SEQ ID 5258>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2266(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:CAB13015 GB:Z99110 yjbK [Bacillus subtilis]
 Identities = 63/184 (34%), Positives = 99/184 (53%), Gaps = 10/184 (5%)

Query: 4 LEIEYKTLNKNDEFNRLTSLFSHVQP--ITQTNYYFDTETFEKHAHRMSLRIRTLNRAE 61
 +EIE+K +L K EF + S + Q N+YFDT++F +K +LRIR +
 25 Sbjct: 5 IEIEFKNMLTKQEFKNIASALQLTEKDFDTQKNHYFDTSFALKQKHAALRIRKNGKYV 64

Query: 62 LTLKIPREVGNGLEHNHDLT--LEEAKYIVKNGQFPEDTEIASLILEKGVDPKLA VFGQL 119
 LTLK P +VG LE + L+ + A + V G P ++ L +D + FG L
 30 Sbjct: 65 LTLKEPADVGLLETHQQLSEVSDLAGFSVPEG--PVKQQLHKL----QIDTDAIQYFGSL 118

Query: 120 TTTTREMETSIGLMALDSNIYADIKDYELELEVQKPKQKGRDFDQFLKENNINFKYAKSK 179
 T R E ET GL+ LD + Y + +DYE+E E +G++ F++ L++ +I + K+K
 35 Sbjct: 119 ATNRAEKETEKGILVLDHSRYLNKEDYEIEFEADWHEGRQAFKLLQQFSIPQRETKNK 178

Query: 180 VARF 183
 + RF
 35 Sbjct: 179 ILRF 182

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5259> which encodes the amino acid sequence <SEQ ID 5260>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.3470(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 114/188 (60%), Positives = 139/188 (73%), Gaps = 1/188 (0%)

Query: 1 MTHLEIEYKTLNKNDEFNRLTSLFSHVQPITQTNYYFDTETFEKHAHRMSLRIRTLNRA 60
 MT+LEIEYKTLN K+E+NRL S HV P+TQTNYY DT+ F++KA++MSLRIRT N A
 55 Sbjct: 1 MTNLEIEYKTLNLTKEYNRLLSQMKHVTPVTQTNYYIDTKAFDLKANKMSLRIRTFVNSA 60

Query: 61 ELTLKIPREVGNGLEHNHDLTLEEAKYIVKNGQFPEDTEIASLILEKGVDPKLA VFGQLT 120
 ELTLK+P +VGN E+N L LE+AK ++K+G PE T + +I+ KG+ P+ L FG LT
 Sbjct: 61 ELTLKVPEKVGNGREYNVPLFLEQAKMDIKHGNLPESTAL-DIIISKGIKPSALVTFGNLT 119

-1908-

Query: 121 TTRREMETSIGLMALDSNIYADIKDYELELEVKQPKQKQKDFDQFLKENNINFKYAKSKV 180
 T RRE IG +ALD N+YA+ KDYELELEV QGK DFD FL E +I FKYAKSKV
 Sbjct: 120 TVRRETVIPIGKLALDYNLYANTKDYELELEVSDALQGKIDFDSFLSEYHITFKYAKSKV 179

5 Query: 181 ARFSATLK 188
 AR TLK
 Sbjct: 180 ARCINTLK 187

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 1694

A DNA sequence (GBSx1798) was identified in *S.agalactiae* <SEQ ID 5261> which encodes the amino acid sequence <SEQ ID 5262>. Analysis of this protein sequence reveals the following:

15 Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1815(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 25 vaccines or diagnostics.

Example 1695

A DNA sequence (GBSx1799) was identified in *S.agalactiae* <SEQ ID 5263> which encodes the amino acid sequence <SEQ ID 5264>. Analysis of this protein sequence reveals the following:

30 Possible site: 38
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0621(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 40 vaccines or diagnostics.

Example 1696

A DNA sequence (GBSx1800) was identified in *S.agalactiae* <SEQ ID 5265> which encodes the amino acid sequence <SEQ ID 5266>. This protein is predicted to be ribose-phosphate pyrophosphokinase (prsA). Analysis of this protein sequence reveals the following:

45 Possible site: 22
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1909-

bacterial cytoplasm --- Certainty=0.3369(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11827 GB:Z99104 phosphoribosyl pyrophosphate synthetase
 [Bacillus subtilis]

Identities = 166/319 (52%), Positives = 231/319 (72%), Gaps = 4/319 (1%)

10 Query: 1 MAEQYADKQIKLFSLTANREIAEKISQASGIPLGKMSSRQFSDEGEIMINIEETVRGDDIY 60
 M+ QY DK +K+FSL +N E+A++I+ G+ LGK S +FSDGE+ INIEE++RG D Y
 Sbjct: 1 MSNQYGDKNLKFSLNSNP ELAKEIADIVGVQLGKCSVTRFSDGEVQINIEESIRGCDY 60

15 Query: 61 IIQSTSFPVNDNLWELLIMIDACKRASANTVNIVVPYFGYSRQDRIAASREPITAKLVAN 120
 IIQSTS FVN+++ ELLIM+DA KRASA T+NIV+PY+GY+RQDR A SREPITAKL AN
 Sbjct: 61 IIQSTSDPVNEHIMELLIMVDALKRASAKTINIVIPYGYARQDRKARSREPITAKLFAN 120

20 Query: 121 MLVKAGVDRVLTLDLHAVQVQGFDDIPVDNLFTVPLFAEHYNQLGLSGEDVVVVSPKNSG 180
 +L AG RV+ LDLHA Q+QGFDDIP+D+L VP+ E++ G + ED+V+VSP + G
 Sbjct: 121 LLETAGATRVIALDLHAPQIQGFDDIPIDHLMGVPILGGEYFE--GKNLEDIVIVSPDHGG 178

25 Query: 181 IKRARS LAEYLDSPAIIDYAQD-DSEREEGYIIGEVEGKKAIIDDIINTGKTFAEAAK 239
 + RAR LA+ L +PIAID + + E I+G +EGK AI+IDDI++T T AA
 Sbjct: 179 VTRARKLADRLKAPIAIDKRRPRPNVAEVMNIVGNIEGKTAILIDDIIDTAGTITLAA 238

30 Query: 299 IADAIIRIHERKPLSPLFS 317
 +A+AIIR+HE++ +S LFS
 Sbjct: 299 LAEAIIRVHEQQSVSYLFS 317

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5267> which encodes the amino acid
 35 sequence <SEQ ID 5268>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.1830(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 278/324 (85%), Positives = 305/324 (93%)

Query: 1 MAEQYADKQIKLFSLTANREIAEKISQASGIPLGKMSSRQFSDEGEIMINIEETVRGDDIY 60
 M E+YADKQIKLFSLT+N IAEKI++A+GIPLGKMSSRQFS+GEIMINIEETVRGDDIY
 50 Sbjct: 1 MTERYADKQIKLFSLTNSNPLFAEKIAAAGIPLGKMSSRQFSNGEIMINIEETVRGDDIY 60

Query: 61 IIQSTSFPVNDNLWELLIMIDACKRASANTVNIVVPYFGYSRQDRIAASREPITAKLVAN 120
 IIQSTSFPVNDNLWELLIMIDACKRASANTVNIV+PYFGYSRQDR+A REPITAKLVAN
 55 Sbjct: 61 IIQSTSFPVNDNLWELLIMIDACKRASANTVNIVLPYFGYSRQDRVAKPREPITAKLVAN 120

Query: 121 MLVKAGVDRVLTLDLHAVQVQGFDDIPVDNLFTVPLFAEHYNQLGLSGEDVVVVSPKNSG 180
 ML KAG+DRV+TLDLHAVQVQGFDDIPVDNLFTVPLFAE Y++LGLSG DVVVVSPKNSG
 60 Sbjct: 121 MLTKAGIDRVVLTLDLHAVQVQGFDDIPVDNLFTVPLFAERYSKLGLSGSDVVVVSPKNSG 180

Query: 181 IKRARS LAEYLDSPAIIDYAQDDSEREEGYIIGEVEGKKAIIDDIINTGKTFAEAAKI 240
 IKRARS LAEYLDSPAIIDYAQDDSEREE+GYIIG+V GKKA I+IDDIINTGKTFAEAAKI
 65 Sbjct: 181 IKRARS LAEYLDSPAIIDYAQDDSEREEGYIIGDVSGKKAILIDDIINTGKTFAEAAKI 240

Query: 241 LERGGATEIYAVASHGLFAGGAADILESAPIREIIVTDSVLSKERIPSNIKYLTASHLIA 300
 LER GAT+ YAVASHGLFAGGAAD+LE+API+EIIVTDSV +K R+P N+ YL+AS LIA

-1910-

Sbjct: 241 LERSGATDTYAVASHGLFAGGAADVLETAPIKEIIVTDSVKTKNRVPENVTYLSASDLIA 300

Query: 301 DAIIRIHERKPLSPLFSYRSDKKD 324
+AIIRIHER+PLSPLFSY+ K+

5 Sbjct: 301 EAIIRIHERRPLSPLFSYQPKGKN 324

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1697

10 A DNA sequence (GBSx1801) was identified in *S.agalactiae* <SEQ ID 5269> which encodes the amino acid sequence <SEQ ID 5270>. This protein is predicted to be Fe-S cluster formation protein. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1981(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

20

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04979 GB:AP001511 Fe-S cluster formation protein [Bacillus halodurans]
Identities = 174/373 (46%), Positives = 237/373 (62%), Gaps = 6/373 (1%)

25

Query: 3 IYLDNAATTALTPSVIEKMTNVMTSNYGNPSSIH TFG RQANQLLRECRQIIAEYLNVNSR 62

IYLD+AAT+ + P VI+ M +GNPSSIH FGR+A Q + E R IA L +

Sbjct: 4 IYLDHAATSPVHPEVIQAMLPYEEQFGNPSSIHQFGRRARQGVDEARGTIARLLQADPS 63

30

Query: 63 EIIFTSGGTESNNTAIKGYALANQLKGKHIITSEIEHHSVLHTMTYLSERFGFDITYLKP 122

E IFTSGGTE++N AI GYA ++ KG HIITS++EHH+VLH L E GF++TY+

Sbjct: 64 EFIFTSGGTEADNLAIFGYAYQHRGKGNHIITSQVEHHAVLHACQEL-EHQGFVITYVPV 122

35

Query: 123 NH-GQITAKDVQEQALRDDTIMVSLMFVNNETGDFLPIQEIGQLLRNHQAVFHVDAVQVFS 181

+ G+++ +DV++ALRDDTI+V+LM+ NNE G PI EIG LL++HQAV H DAVQ F

Sbjct: 123 DQGRVSVEDVRQALRDDTILVTLMYGNNEVGTIQPIAEIGALLQDHQAVLHTDAVQAFG 182

40

Query: 182 KMELDPHSLGIDFLAASAHKFHGPKGVGILYCAPH-HFDSLHGGDQEEKRRASTENIIG 240

+ ++ L +D L+ SAHK +GPKGVG+LY L+GG+QE K+RA TEN+

Sbjct: 183 AISIELDHLFVDMLSVSAHKINGPKGVGLLYVRDGIIVLKPALYGGEQERKKRAGTENVAA 242

45

Query: 241 IAGMSQALTDATTNLKNWTHISQLRTTFLDAISD--LDFYLNNGQDC-LPHVLNIGFPG 297

I G ++A+ A N + TF D + F +N Q LPH+ N+ FPG

Sbjct: 243 IIGFAKAVEIAIANREERQKAYFDYCQTFDQFQEGVQFVMNGHQTWRLPHIFNVSFPG 302

50

Query: 298 QNNGLLLTQLDLAGFAVSTGSACTAGTVEPSHVLTSLYGANSPLNSESIRISFSELNTQE 357

+ LL LDLAG A S+GSACTAG++EPSHVL +++G++S + +R SF NT+E

Sbjct: 303 VHVEALLVNLDLAGIAASSGSACTAGSIEPSHVLVAMHGSSELVTSGVRFSFGLGNTKE 362

Query: 358 EILELAKTLRKII 370

+ AK KI+

Sbjct: 363 HVQWAAKETAKIV 375

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5271> which encodes the amino acid sequence <SEQ ID 5272>. Analysis of this protein sequence reveals the following:

55

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1477(Affirmative) < succ>

-1911-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 235/370 (63%), Positives = 285/370 (76%)

Query: 2 MIYLDNAATTALTSPVIEKMTNVMTSNYGNPSSIH FGRQANQLLRECRQIIAEYLVNS 61
 M Y DNAATT L+P+VI MT M N+GNPSSIH +GR+AN++LRECRQ IA L +
 10 Sbjct: 1 MTYFDNAATTPLSPNVIRAMTAAMQDNFGNPSSIH FYGRRANKILRECRQAIARNLGASE 60

Query: 62 REIIFTSGGTESNNTAIKGYALANQLKGKHIITSEIEHHSVLHTMTYLSERFGFDITYLK 121
 ++II TSGGTESNN AIKGYALA+Q KGKH+IT+ IEHHSVLHTM YL ERFGE++TYL
 15 Sbjct: 61 QQIIVTSGGTESNNMAIKGYALAHQAKGKHLITTTIEHHSVLHTMAYLEERFGFEVITYLP 120

Query: 122 PNHGQITAKDVQEALRDDTIMVSLMFVNNETGDFLPIQEIGQLLRNHQAVFHVDAVQVFS 181
 +GQI D+++ALRDDTI+VS+M+ NNETGD LPI++IG LL++HQA FHVDAVQ
 15 Sbjct: 121 CQNGQINLSDLKQALRDDTILVSIMYANNETGDLLEPIKDIGNLLKDHQAFAFHVDAVQAVG 180

Query: 182 KMELDPHSLGIDFLAASAHKFHGPKGVGILYCAPHHFDSLHGGDQEEKRRASTENIIGI 241
 K+++ P LGIDFL+ASAHKFHGPKG G LY D LLHGGDQE KRRASTEN++GI
 20 Sbjct: 181 KLKIIPSELGIDFLSASAHKFHGPKGCGFLYSNGQPIDPLHGGDQEGKRRASTENMLGI 240

Query: 242 AGMSQALTDATINTLKNWTHISQLRTTFLDAISDLDFYLNNGQDCLPHVLNIGFPGQNNG 301
 GM+QALTD A T ++ HI LR + + L +Y+N G LPHVLNIGF G N
 25 Sbjct: 241 IGMAQALTDAMTCLDQSTDHIIISLRHHLISLLEGLPYIYINQGTHYLPHVLNIGFLGYQNT 300

Query: 302 LLLTQLDLAGFAVSTGSACTAGTVEPSHVLTSLYGANSPRINESIRISFSELNTQEEILE 361
 +LLTQLDLAG AVSTGSACTAG V PSHVL + YG +S RL ESIRISFS+ N+ E++ +
 30 Sbjct: 301 ILLTQLDLAGIAVSTGSACTAGAVNPSHVLAAYYGDDSSRLKESIRISFSQNSIEDVNQ 360

Query: 362 LAKTLRKIIIG 371
 LA+TL+ I+G
 Sbjct: 361 LAQTLKNILG 370

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1698

A DNA sequence (GBSx1802) was identified in *S.agalactiae* <SEQ ID 5273> which encodes the amino acid sequence <SEQ ID 5274>. Analysis of this protein sequence reveals the following:

40 Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.2753 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12416 GB:Z99107 ydiH [Bacillus subtilis]
 50 Identities = 96/202 (47%), Positives = 140/202 (68%), Gaps = 4/202 (1%)

Query: 7 IPKATAKRLSLYYRIFKRFNTDGIKASSKQIADALGIDSATVRRDFS YFGELGRRGFGY 66
 IP+ATAKRL LYYR K + G ++ SS +++DA+ +DSAT+RRDFS YFG LG++G+GY
 55 Sbjct: 8 IPQATAKRLPLYYRFLKNLHAGSKQVRSSAELSDAVKVD SATIRRRDFS YFGALGKKGYGY 67

Query: 67 DVKKLMNFFAEILNDHSTTNVMLVGCNIGRALLHYRFHNRNMQISMAFDLSDNLDLVGK 126
 +V L+++FF + L+ T+V+L+G GN+G A LHY F N +ISMAFD++ + +
 Sbjct: 68 NVDYLLSFFRKTLDDQEMTDVILIGVGNLGTAF LHYNFTKNNTKISMAFDINESKI--G 125

60 Query: 127 TTEDGIPVYGISTINDHLIDSDIETAILTVPSTEAQEVADILVKAGIKGILSFSPVHLTL 186
 T G+PVY + + H+ D + AILTVP+ AQ + D LV GIKGIL+F+P L +

-1912-

Sbjct: 126 TEVGGVPVYNLDDLEQHVKDESV--AILTVPAVAQAQSI'DRLVALGIKILNFTPARLNV 183

Query: 187 PKDIIQYVDLTSELQTLLEYFM 208

P+ I + ++DL ELQ+L+YF+

Sbjct: 184 PEHIRIHHIDLAVELQSLVYFL 205

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5275> which encodes the amino acid sequence <SEQ ID 5276>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2313(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 167/210 (79%), Positives = 189/210 (89%)

Query: 1 MIMDKSIPKATAKRLSLYYRIFKRFNTDGIKASSKQIADALGIDSATVRRDFSYPGELG 60

+++DKSIPKATAKRLSLYYRIFKRF+ D +EKASSKQIADA+GIDSATVRRDFSYPGELG

Sbjct: 1 VVIDKSIPKATAKRLSLYYRIFKRFHADQVEKASSKQIADAMGIDSATVRRDFSYPGELG 60

Query: 61 RRGFGYDVVKLMNFFAEILNDHSTTNVMLVGCNIGRALLHYRFHNRNMQISMAFDLDS 120

RRGFGYDV KLMNFFA++LNDHSTTNV+LVGCNIGRALLHYRFHNRNMQI+M FD D

Sbjct: 61 RRGFGYDVTKLMNFFADLLNDHSTTNVILVGCNIGRALLHYRFHNRNMQIAMGFDTD 120

Query: 121 NDLVGKTTEDGIPVYGISTINDHLIDSDIETAILTVPSTEAQEVADILVKAGIKGILSFS 180

N LVG T D IPV+GIS++ + + ++DIETAILTVPS AQEV D L++AGIKGILSF+

Sbjct: 121 NALVGTKTADNIPVHGISSVKERIANTDIETAILTVPSIHAQEVTDQLIEAGIKGILSFA 180

Query: 181 PVHLTLPKDIIQYVDLTSELQTLLEYFMNQ 210

PVHL +PK +IVQ VDLTSELQTLLEYFMNQ

Sbjct: 181 PVHLQVPKGIVQSVDLTSELQTLLEYFMNQ 210

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1699

A DNA sequence (GBSx1803) was identified in *S.agalactiae* <SEQ ID 5277> which encodes the amino acid sequence <SEQ ID 5278>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2966(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9847> which encodes amino acid sequence <SEQ ID 9848> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14764 GB:Z99118 similar to DNA repair protein [Bacillus subtilis]

Identities = 90/210 (42%), Positives = 136/210 (63%)

Query: 24 PRERLVDLGADRLSNQELLAILLRTGIKEKPVLEISTQILENISSLAIDFGQLSLQELQSI 83

PRERL+ +GA+ L+N ELLAILLRTG K + VL++S ++L + L + S++EL SI

Sbjct: 19 PRERLLKVGAENLANHELLAILLRTGTHESVLDLSNRLRLRSFDGLRLLKEASVEELSSI 78

-1913-

Query: 84 KGIGQVKSVEIKAMLELAKRIHKA EYDRKEQILSSEQLARKMMLELGDKKQEH LVAIYMD 143
 GIG VK+++I A +EL RIHK + I S E A +M ++ QEH V +Y++
 Sbjct: 79 PGIGMVKAIQILA AVELGSR I HKLANEEHFVIRSPEDGANLV MEDMRFLTQEH FVCLYLN 138

Query: 144 TQNR IIEQRTIFIGTVRRSVAEPREILHYACKNMATSL I IHNHPSGSPKPS ESDL SFTK 203
 T+N++I +RT+FIG++ S+ PRE+ A K A S I +HNHPSG P PS D+ T+
 Sbjct: 139 TKNQVIHKRTVFIGSLNSSIVHPREVFKEAFKRS AASFICVHNHPSG DPTPSREDIEVTR 198

Query: 204 KIKRSCDHLGIVCLDHIIVGKNKYYSFREE 233
 ++ + +GI LDH+++G K+ S +E+
 Sbjct: 199 RLFECGNLIGIELLDHLVIGDKKFVSLKEK 228

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5279> which encodes the amino acid sequence <SEQ ID 5280>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3307(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 145/225 (64%), Positives = 182/225 (80%)

Query: 12 MYHIELKKEALLPRERLVDL GADRLSNQELLAILLRTGIKEKPVLEISTQILENISSLAD 71
 MY I+ +PRERL+ LGA+ LSNQELLAILLRTG KEK VLE+S+ +L ++ SLAD
 Sbjct: 1 MYSIKCDNKKAMPRELRMLGABSLSNQELLAILLRTGNKEKHVLELSSYLLSHLDSLAD 60

Query: 72 FGQLSLQELQSIKGIGQVKSVEIKAMLELAKRIHKA EYDRKEQILSSEQLARKMMLELG D 131
 F ++SLQELQ + GIG+VK++EIKAM+EL RI + + +L+S Q+A KMM LGD
 Sbjct: 61 FKKMSLQELQHLAGIGKVKAI EIKAMIELVSRILATDKTLLTDSVLT SVQVAEKMMALGD 120

Query: 132 KKQEH LVAIYMDTQNR IIEQRTIFIGTVRRSVAEPREILHYACKNMATSL I IHNHPSGS 191
 KKQEH LV +Y+D QNR IIE++TIFIGTVRRS+AEPREIL+YACKNMATSLI+IHNHPSG+
 Sbjct: 121 KKQEH LVVLYLDNQNR IIEKTIFIGTVRRSLAEPREILY YACKNMATSLIV IHNHPSGN 180

Query: 192 PKPSES DLSFTK KIKRSCDHLGIVCLDHIIVGKNKYYSFREEADI 236
 +PS +D FT+KIKRSC+ LGI+CLDHIIV YYSFRE++ +
 Sbjct: 181 IEPSSNDYCFTEKIKRSCEDLGIICLDHIIVSYKDYYSFREKSTL 225

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1700

A DNA sequence (GBSx1804) was identified in *S.agalactiae* <SEQ ID 5281> which encodes the amino acid sequence <SEQ ID 5282>. This protein is predicted to be a permease. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -7.86	Transmembrane	258 - 274 (255 - 290)
INTEGRAL	Likelihood = -7.32	Transmembrane	89 - 105 (79 - 109)
INTEGRAL	Likelihood = -4.88	Transmembrane	176 - 192 (170 - 194)
INTEGRAL	Likelihood = -4.78	Transmembrane	339 - 355 (326 - 359)
INTEGRAL	Likelihood = -4.57	Transmembrane	237 - 253 (236 - 257)
INTEGRAL	Likelihood = -3.98	Transmembrane	39 - 55 (38 - 59)
INTEGRAL	Likelihood = -3.40	Transmembrane	292 - 308 (282 - 308)
INTEGRAL	Likelihood = -1.38	Transmembrane	317 - 333 (317 - 333)
INTEGRAL	Likelihood = -0.27	Transmembrane	8 - 24 (8 - 24)

-1914-

----- Final Results -----

bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC05771 GB:AF051356 putative permease [Streptococcus mutans]

Identities = 88/366 (24%), Positives = 175/366 (47%), Gaps = 27/366 (7%)

Query: 3 FEKRQVYVVITFAICYAIQAYW---GAVSNILTTLHKAIF-PFLMGAGIAYIINIVMSV 58
F+ ++++ + + I W G++ N ++ K F PFL+G + YI N +++
Sbjct: 2 FKSSKLFFWTVEILLVTLILFIWRQMSIFNPFSSVAKTFFLPFLGGFLYYITNPIVTF 61

Query: 59 YERLYIKLFGKSRLMAIKRSVSMILSYATFIGLIVWLFISIVIPDLISSLSLLVIDTGA 118
E + IKR + L +A + L+V+ + +IP+LI+ L+ L+
Sbjct: 62 LENRF-----KIKRIWGITLIFAVLLSLLVFSITSLIPNLINQLTDLISASQNI 110

Query: 119 LAKLVNNLNENKQISEVLNMGTDKDLVSTLSGYSQQILKQVLSVLTNLLTSVSSIAATL 178
L + NE K N D+ L ++ + + +VL ++ SVSSI +
Sbjct: 111 YVGLQDLFNWKSNPAPFKNI-----DIPVLLKQFNLSYVDILTIVLDSVTVSIVYMI 165

Query: 179 LNVFVSFIFS----IYVLANKQLGRQFNLLIDTYLGSTGKTFHYVRHILHQRFHGFFVS 234
N + + + Y+L +K+ L +L T L + + + + +
Sbjct: 166 TNTVMILVLTTPVILFYLLKDKDGL--MPMLDRTILKNDRHNISQLLNQMNKTISRYSIS 222

Query: 235 QTLEAMILGSLTVIGMLIFQFPYALTGVGLVAFTALIPVVGAYIGVTIGFILIATESLTE 294
++A + +IG I YA ++ T +IP VG Y+G+T + +
Sbjct: 223 VAIDAAFIFVFALIGYQIMGVQYAFALFALVAGITNVIYPVGPYGLTPVVLAYVVSPPK 282

Query: 295 AFLFVLFLILLQQFEGNVIYPKVVGSGILPSMWVLMAITIGGALWGILGMLLAVPVAAT 354
+ +++++ LQQ +GN++YP+VVG ++ + + + + +GG + G++GML+AVP A
Sbjct: 283 MIIAIIYIMTLQQIDGNIVYPRVVGSTMKIHLPTIMVLLVLGGNIAGLVGMLVAVPAYAI 342

Query: 355 IYQIVK 360
I +IVK
Sbjct: 343 IKEIVK 348

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5283> which encodes the amino acid sequence <SEQ ID 5284>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -8.70	Transmembrane	87 - 103 (83 - 116)
INTEGRAL	Likelihood = -7.27	Transmembrane	178 - 194 (166 - 202)
INTEGRAL	Likelihood = -6.74	Transmembrane	278 - 294 (256 - 297)
INTEGRAL	Likelihood = -5.41	Transmembrane	299 - 315 (295 - 321)
INTEGRAL	Likelihood = -4.46	Transmembrane	14 - 30 (13 - 32)
INTEGRAL	Likelihood = -3.56	Transmembrane	340 - 356 (333 - 366)
INTEGRAL	Likelihood = -3.35	Transmembrane	258 - 274 (256 - 277)

----- Final Results -----

bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC05771 GB:AF051356 putative permease [Streptococcus mutans]

Identities = 87/373 (23%), Positives = 168/373 (44%), Gaps = 41/373 (10%)

Query: 10 FEKKQVFLVLTFLICYGILANWRNGTAIVTTIYKTS---LPFFYGAGAYIVNIVMSA 65
F+ ++F+ + +L IL WR +I + + LPF G YI N +++
Sbjct: 2 FKSSKLFFWTVEILLVTLILFIWRQMSIFNPFSSVAKTFFLPFLGGFLYYITNPIVTF 61

Query: 66 YEKVYVYIFKDWSHVLKVKRGICLLLAYLTFILITWIIISIVIPDLITSISTLTKFDT-- 123

-1915-

E + K+KR + L + L+ + I+ +IP+LI ++ L
 Sbjct: 62 LENRF-----KIKRIWGITLIFAVLLSLLVFSITSLIPNLINQLTDLISASQNI 110
 Query: 124 -ITIQEVVNNLEHNKLLARTIQYIGGDGKLTETIANYSQQLLKQFLTVLTNLTSTVTIA 182
 + +Q++ N + N I +Q ++ +LTN+L SVTV
 Sbjct: 111 YVGLQDLFNEWKSNPAFKNI-----DIPVLLKQFNLSYVDILTNVLDSTVSV 158
 Query: 183 SAIINLFISFVPSL-----YVLASKEDLCRQGNLTVDYTTGKYAKRIHYLLELLHQR 234
 S+I+ + + V L Y+L K+ L L T I LL +++
 Sbjct: 159 SSIVYMITNTVMILVLTVPVILFYLLKDKDGLMPM---LDRTILKNDRHNSQLLNQMNKT 215
 Query: 235 FHGFFVSQTL EAMILGSLTASGMFILRLPFAGTIGVLVAFTALIPVIGASIGAAIGFILI 294
 + ++A + G I+ + +A ++ T +IP +G +G +
 Sbjct: 216 ISRYISGVAIDAAFIFVFALIGYQIMGVQYAFLFALVAGITNVIPYVGPYLG LTPVVLAY 275
 Query: 295 MTQSMSQAIIFIIFLIILQQIEGNFIYPKVVGSGIGLPAMWVLMAITIGASLKGIVGMII 354
 + + II II+++ LQQI+GN +YP+VVG ++ + + +++ + +G ++ G+VGM++
 Sbjct: 276 VVSDPKMIIAIYIMTLQQIDGNIVYPRVVGSTMKIHLPTIMVLLVLGGNIAGLVGMLV 335
 Query: 355 AVPLAATLYQVIK 367
 AVP A + +++K
 Sbjct: 336 AVPAYAIKEIVK 348

An alignment of the GAS and GBS proteins is shown below.

Identities = 218/370 (58%), Positives = 291/370 (77%)
 Query: 1 MKFEKRQVYVYVITFAICYAIQAYWGAVSNILTTLHKAIFPFLMGAGIAYIINIVMSVYE 60
 MKFEK+QV+Y+V+TF +CY I A W + I+TT++K PF GA AYI+NIVMS YE
 Sbjct: 8 MKFEKKQVFYLVLTFILCYGILANWRNGTAIVTTIYKTSLPFFYGAAGAYIVNIVMSAYE 67
 Query: 61 RLYIKLFKGSRLMAIKRSVSMILSYATFIGLIVWLF SIVIPDLISSLSLLVIDTGALA 120
 ++Y+ +FK ++ +KR + ++L+Y TF LI W+ SIVIPDLI+S+S+L DT +
 Sbjct: 68 KVVYVIFKDWSHVLKVKRGICLLAYLTFFILITWIIISIVIPDLITSISTLTKEFTITIQ 127
 Query: 121 KLVNNLNENKQISEVLNMGTDKDLVSTLSGYSQQILKQVLSVLTNLLTSVSSIAATLLN 180
 ++VNNL NK ++ + Y+G D L T++ YSQQ+LKQ L+VLTN+LTSV+ IA+ ++N
 Sbjct: 128 EVVNNLEHNKLLARTIQYIGGDGKLTETIANYSQQLLKQFLTVLTNLTSTVTIASAIIN 187
 Query: 181 VFVSFIFSIYVLANKQLGRQFNLLIDTYLGSTGKTFHYVRHILHQRFHGFFVSQTL EAM 240
 +F+SF+FS+YVLA+KE L RQ N L+DTY G K HY+ +LHQRPHGFFVSQTL EAM
 Sbjct: 188 LFISFVFSLYVLASKEDLCRQGNLTVDYTTGKYAKRIHYLLELLHQRFHGFFVSQTL EAM 247
 Query: 241 ILGSLTVIGMLIFQFPYALTGVGLVAFTALIPVVGAYIGVTIGFILIATESL TEAFLFVL 300
 ILGSLT GM I + P+A T+GVLVAFTALIPV+GA IG IGFI LI T+S+++A +F++
 Sbjct: 248 ILGSLTASGMFILRLPFAGTIGVLVAFTALIPVIGASIGAAIGFILINTQSMSQAIIFII 307
 Query: 301 FLIILQQFEGNVIYPKVVGSGIGLPMSMWVLMAITIGGALWGILGMLLAVPVAATYQIVK 360
 FLI+LQQ EGN IYPKVVGSGIGLP+MWVLMAITIG +L GI+GM++AVP+AAT+YQ++K
 Sbjct: 308 FLIILQQIEGNFIYPKVVGSGIGLPAMWVLMAITIGASLKGIVGMIIAVPLAATLYQVIK 367
 Query: 361 DHIKRQTLR 370
 D+I KRQ ++
 Sbjct: 368 DNIQRQAIQ 377

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1701

A DNA sequence (GBSx1805) was identified in *S.agalactiae* <SEQ ID 5285> which encodes the amino acid sequence <SEQ ID 5286>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence

-1916-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1081(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

A related GBS nucleic acid sequence <SEQ ID 9849> which encodes amino acid sequence <SEQ ID 9850> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAA69226 GB:U29579 6-phospho-beta-glucosidase [Escherichia coli]
 Identities = 290/478 (60%), Positives = 369/478 (76%), Gaps = 2/478 (0%)

Query: 2 MVKQVFPKGFLWGGATAANQCEGAYNVDRGLANVDVVPVTGEDRFALISGQKKMFDFEEG 61
 M VFP+ FLWGA AANQ EGA+ +GL VD++P GE R A+ G +K F +
 Sbjct: 1 MKMSVFPESFLWGGALAAQSEGAFFREGDKGLTTVDMIPHGEHRMAVKLGLEKRFQLRDD 60

15 Query: 62 YFYPAKESIDFYHHYKEDLALLAEMGFKTYRMSIAWTRIFPKGDELYPNEAGLQFYENIF 121
 FYP+ E+ DFYH YKED+AL+AEMGFK +R SIAW+R+FP+GDE+ PN+ G+ FY ++F
 Sbjct: 61 EFYPSHEATDFYHRYKEDIALMAEMGFKVFRTSIAWSRLFPQGDITPNQOGIAFYRSVF 120

20 Query: 122 KECRKYGIEPLVTITTHFCPIYLIKHYGGWRSRKMIGFYERLVRALFTRFKGLVKYWLTF 181
 +EC+KYGIEPLVT+ HFD P++L+ YG WR+RK++ F+ R R F F GLVKYWLTF
 Sbjct: 121 EECKKYGIEPLVTLCHFDPVPHLVTEYGSWRNRKLVEFFSRYARTCFEAFDGLVKYWLTF 180

25 Query: 182 NEINMILHAPFMGAGLYFEDGENQEIKYQAAHHELVASAIAVKIAHEVDPNNQIGCMLA 241
 NEIN++LH+PF GAGL FE+GENQ+Q+KYQAAHH+LVASA+A KIAHEV+P NQ+GCMLA
 Sbjct: 181 NEINIMLHSPFSGAGLVPEEGNQDQVKYQAAHHQLVASALATKIAHEVNPQNQVGCMLA 240

30 Query: 242 AGQYYPNTCHPQDYWASMOKNRENYFFIDVQARGKYPNYAKKHFEHLGISIQMTAEDLAL 301
 G +YP +C P+D WA+++K+REN FFIDVQARG YP Y+ + F G++I D +
 Sbjct: 241 GGNFYFYSCKPEDVWAALKDRENLFIDVQARGTYPAYSARVFREKGVITINKAPGDDEI 300

35 Query: 302 LRDTYVDFISFSYSSRVASGNPTVSEQVQENIFASLKNPYLKSSEWGWQIDPLGLRITL 361
 L++ TVDF+SFSYY+SR AS + N+ SL+NPYL+ S+WGW IDPLGLRIT+
 Sbjct: 301 LKN-TVDFVSFSYASRCSAEMNANSSAANVVKSLRNPYLQVSDWGWGIDPLGLRITM 359

40 Query: 362 NAIWDYQKPMFIVENGLGAVDIPDENGVEDDYRIDYLRQHIAAMRDAIYVDGVNLIGY 421
 N ++DRYQKP+F+VENGLGA D NG + DDYRI YLR+HI AM +AI DG+ L+GY
 Sbjct: 360 NMMYDRYQKPLFLVENGLGAKDEFAANGEINDDYRISYLRHIRAMGEAI-ADGIPLMGY 418

Query: 422 TTWGCIDLVSAGTGEMEKRYGFIYVDRNNKGEGLTKRYKKKSFYWKVVIASNGSQIE 479
 TTWGCIDLVA TGEM KRYGF++VDR++ G GTL R +KKSF+WYKKVIASNG +E
 Sbjct: 419 TTWGCIDLVSASTGEMSKRYGFVVDREDDAGNGLTRTRKKSFWYKKVIASNGEDLE 476

There is also homology to SEQ ID 5288.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1702

50 A DNA sequence (GBSx1806) was identified in *S.agalactiae* <SEQ ID 5289> which encodes the amino acid sequence <SEQ ID 5290>. This protein is predicted to be platelet-activating factor acetylhydrolase isoform Ib beta subunit, pu. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.5323(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1917-

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC27974 GB:AF016048 platelet-activating factor acetylhydrolase
alpha 2 subunit [Rattus norvegicus]
Identities = 43/177 (24%), Positives = 84/177 (47%), Gaps = 9/177 (5%)

5 Query: 28 QEGAIVFTGDSIVEF---FPLKKHLGRDYPLVNRGVAGSDTYWLLLENLRTQVWELLPSKV 84
+E ++F GDS+V+ + + + L +N G+ G T +L L+ E + KV
Sbjct: 38 KEPDVLVFGDSMVQLMQQYEIWRLEFSPHLALNFGIGGDTTRHVLWRLKNGELENIKPKV 97

10 Query: 85 FIL-IGTNDIGLGHQSQSEIIANITDIIAEIRAESYMTINILSVLPVSEEDDYIERVKVR 143
++ +GTN+ ++ E+ I I+ I +I +L +LP E+ + + + +
Sbjct: 98 IVVWVGTTNNHE--NTAAEVAGGIEAIVQLINTRQPOAKIIVLGLLPGEKPNPLRQKNK 155

15 Query: 144 NNQTIKALNKTLVSGINIEYDILLVDEKGQLASSFTKDGLHLTDQAYAKISETI 200
NQ +K +L ++ + +++ V G ++ D LHLT YAKI + +
Sbjct: 156 VNQLLKV---SLPKLANVQLLDIDGGFVHSDGAISCHDMFDLHLTGGGYAKICKPL 209

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5291> which encodes the amino acid sequence <SEQ ID 5292>. Analysis of this protein sequence reveals the following:

20 Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.5979(Affirmative) < succ>
25 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 92/204 (45%), Positives = 133/204 (65%)

30 Query: 1 MLEVIDKALRDYQMKREQFFEINNQTVOEGAIVFTGDSIVEFFPLKKHLGRDYPLVNRGV 60
MLE++ + LR YQ ++ + NQ +G IVF GDS++EFFPLKK G P++NRG+
Sbjct: 1 MLEIVSEELRHYQEQLIEYRNKNQLAPKGGIVFAGDSLIEFFPLKKAFGSCLPIINRGI 60

35 Query: 61 AGSDTYWLLLENLRTQVWELLPSKVFILIGTNDIGLGHQSQSEIIANITDIIAEIRAESYMT 120
AG D+ WLL + Q+ +L P +F+LIG NDIGLG+ + I+ I ++I++IR+ +
Sbjct: 61 AGIDSQWLLRHFSVQITDLEPKHIFLLIGCNDIGLGYDKCHIVKTIVELISQIRSHCVYS 120

40 Query: 121 EINILSVLPVSEEDDYIERVKVRNNTIKALNKTLVSGINIEYDILLVDEKGQLASS 180
+I +LS+LPVS Y + VK+R N I A+NK L++I . + +I L L DEKG L+
Sbjct: 121 QIYLLSLLPVSNPNRYQKTVKIRTNAMIDAINKDLAMIPTVEFINLNTCLKDEKGGLSDE 180

45 Query: 181 FTKDGLHLTDQAYAKISETIKLYL 204
T DGLHL AYAK++E IK Y+
Sbjct: 181 NTL DGLHLNFPAYAKLAETIKSYI 204

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1703

50 A DNA sequence (GBSx1807) was identified in *S.agalactiae* <SEQ ID 5293> which encodes the amino acid sequence <SEQ ID 5294>. Analysis of this protein sequence reveals the following:

Possible site: 43
>>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.5226(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1918-

A related GBS nucleic acid sequence <SEQ ID 9851> which encodes amino acid sequence <SEQ ID 9852> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:BAA35556 GB:D90723 Hypothetical 30.2 kd protein in idh-deoR
    intergenic region. [Escherichia coli]
    Identities = 104/265 (39%), Positives = 154/265 (57%), Gaps = 4/265 (1%)

Query: 2  IKLIATDMDGTFLRSDFYDKARFSSLLTMEKYDIKFVAASGNLYDQLLLNFLEYPNRI 61
10  Sbjct: 4  IKLIA DMDGTFL  KTY++ RF + M+ I+FV ASGN Y QL+ F E N I
    IKLIAVDMGTFLSDQKTYNRERFMAQQMKAQGIRFVVASGNQYYQLISFFPEIANEI 63

Query: 62  AYVAENGRVIDQDGTLLKETYLSNDTVAAVLSYLYQNYPETLISLSGEKRSYLERRTPI 121
    A+VAENGG V+ + G + LS D A V+ +L PE I G+ +Y ++
15  Sbjct: 64  AFVAENGWVWVSE-GKDVFNELSKDAFATVVEHLLTR-PEVEIIACGKNSAYTLKKYDD 121

Query: 122  NRRTELEYMPNFIYKDHLLPLDDDRYFQMTLWVNENLVSEMLLDISEHFKNNHIRTSS 181
    +T E Y Y D+ L+D +F+ L +++ L+ ++ + E + + + +
20  Sbjct: 122  AMKTVAEMYHRLEYVDNFDNLEDI-FKFGLNLSDELIPQVQKALHEAIGDIMVSV-HT 179

Query: 182  GFGCIDVLPADVKNADGIAILLEKWLKQDQVMVFGDGGNDVEMLRAANISYAMNAPEE 241
    G G ID++ V+KA+G+ L + WG+ +V+VFGDGGND+EMLR A S+AM NA
25  Sbjct: 180  GNGSIDLIIPGVHKANGLRQLQKLWGIDDSEVVVFGDGGNDIEMLRQAGFSFAMENAGSA 239

Query: 242  IKAIACYQTVSNDQDGVLETIENFL 266
    + A AKY+ SN+++GVL+ I+ L
30  Sbjct: 240  VVAAAKYRAGSNNREGVLDVIDKVL 264

```

There is also homology to SEQ ID 1158.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1704

A DNA sequence (GBSx1808) was identified in *S.agalactiae* <SEQ ID 5295> which encodes the amino acid sequence <SEQ ID 5296>. This protein is predicted to be transcriptional regulator (AraC/XylSfamily). Analysis of this protein sequence reveals the following:

```

35  Possible site: 50
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.4984(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

45  >GP:AAF89977 GB:AF206272 transcriptional regulator [Streptococcus mutans]
    Identities = 195/287 (67%), Positives = 237/287 (81%)

Query: 5  DNLLSHNLEDNRHLLPYEHMTEVRNGYPDILFHHWPELEISYVHEGTARYHIDYDFNS 64
    D H + + LLPY+ T + NGYPD LFHWHPELEISY++EGTA+YHIDYD+FNS
50  Sbjct: 10  DENFKHEINFNDNLLPYKIYQTTIANGYPDTLFFHWHPELEISYIEGTAQYHIDYDYFNS 69

Query: 65  QSGDIILIRPNGMHSHIPIENKEHITDSIKFHLDLIGYSIVDQVSLRYLQPLQTSSFKFI 124
    Q+ DIIL+RENGMHSHIPI+K + ++ FHLDL+GYS++DQ+SLRYLQPLQ S+FK +
55  Sbjct: 70  QTDDIILVRENGMHSHIPIKKNMQKAQTLFLFHLDLVGYSLLDQISLRYLQPLQNSTFKLV 129

Query: 125  QCIKPSMTGYNDIKNCLFDIFNISKEENRHFELLLKAKLNELLYLLYHQQYVIKKHTDDT 184
    CIKP M GY DIKNCLF IF+I + + RHFELLLKAKL EL+YLLY+HQYV++KH+DD
60  Sbjct: 130  PCIKPDMLGYQDIKNCLFAIFDIYQQRGRHFELLLKAKLQELIYLLYFHQYVLRKHSDDM 189

```


-1919-

Query: 185 YRKNERIRDLIDYINNYYQONLTIEFLADYMGYSKTHFMTVFKQHTGTSCTEFIIQVRLN 244
 YRKNE+IR+LIDYI+ +YQ+ L+I LAD +GYSKTHFMTVFKQHTGTSCT+FI IQ RL+
 Sbjct: 190 YRKNEKIRELIDYIHQHYQEKLSIISLADIIGYSKTHFMTVFKQHTGTSCTDFIIQFRLS 249

Query: 245 KASEHLINSTTAIIDIANSVGFNNLSNFRQFKRYHTTPRQYRKQF 291
 KA + L+NS I+++A+ VGF NLSNFRQFKRY TP QYRKQF
 Sbjct: 250 KACDLLVNSIKPILEVASEVGFNLSNFRQFKRYQITPSQYRKQF 296

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5297> which encodes the amino acid sequence <SEQ ID 5298>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 43/169 (25%), Positives = 83/169 (48%), Gaps = 16/169 (9%)

- Query: 136 DIKNCLFDIFNISKEENRHFELLLKAKLNELLYLLYYHQYV-----IKKHTDDTYRKN- 188
 D+K+ F +F+ + R F +L K ++ ++ Q + +KK D T + N
 Sbjct: 319 DVKHSVFLIFS---DIYRQFPILDKMTYLSMVKTIHDSQSIDCILRELKKVLDVTNQNS 375
- Query: 189 -----ERIRDLIDYINNYYQONLTIEFLADYMGYSKTHFMTVFKQHTGTSCTEFIIQVR 242
 + + + ID I Y Q LT++ +AD + + + FK T S T+++ VR
 Sbjct: 376 PEKRYSDLVSETIDCIRKEYHQELTLKAIADRLHVNGVYLGQCFKNETERSFTQYLNHVR 435
- Query: 243 LNKASEHLINSTTAIIDIANSVGFNNLSNFRQFKRYHTTPRQYRKQF 291
 + KA + L+ + +I +IA G+N F + FK+ +P+++R ++
 Sbjct: 436 IQKAQQLLLYTQNSINETAYETGYNTNHYFIKMFKKLNLSPKEFRDRY 484

- 35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1705

A DNA sequence (GBSx1809) was identified in *S.agalactiae* <SEQ ID 5299> which encodes the amino acid sequence <SEQ ID 5300>. Analysis of this protein sequence reveals the following:

- 40 Possible site: 34
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
 bacterial cytoplasm --- Certainty=0.3705(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1920-

Example 1706

A DNA sequence (GBSx1810) was identified in *S.agalactiae* <SEQ ID 5301> which encodes the amino acid sequence <SEQ ID 5302>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -11.25    Transmembrane    59 - 75 ( 56 - 82)
    INTEGRAL    Likelihood = -7.48     Transmembrane    23 - 39 ( 12 - 41)
    INTEGRAL    Likelihood = -6.64     Transmembrane   231 - 247 ( 225 - 255)
    INTEGRAL    Likelihood = -5.15     Transmembrane   335 - 351 ( 333 - 355)
10  INTEGRAL    Likelihood = -4.19     Transmembrane   309 - 325 ( 305 - 327)
    INTEGRAL    Likelihood = -4.14     Transmembrane   272 - 288 ( 268 - 292)
    INTEGRAL    Likelihood = -4.04     Transmembrane   402 - 418 ( 400 - 419)
    INTEGRAL    Likelihood = -3.88     Transmembrane   191 - 207 ( 190 - 208)
    INTEGRAL    Likelihood = -2.71     Transmembrane   365 - 381 ( 364 - 381)
15  INTEGRAL    Likelihood = -1.86     Transmembrane   165 - 181 ( 164 - 182)

----- Final Results -----
        bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF96429 GB:AE004383 conserved hypothetical protein [Vibrio cholerae]
Identities = 142/443 (32%), Positives = 241/443 (54%), Gaps = 20/443 (4%)
25  Query: 6  NEFQFSLESILGFVVRGIVVGLIAGFVVSIFRLAIEKIFLVVMELYKS--AHYQPIILLS 63
    N+F ++ ++ ++VG++AG V + F A+ + + KS + P+ L +
    Sbjct: 21 NQFLSKDKTPFSLVFLSLVLGILAGLVGTIFYEQAVHLVSETRTDWLKSEIGSFLPLWLAA 80

30  Query: 64 ITVTSIIAAVIIGFFI--KSPDPIDKSGGIPHVEGELKGMSPDWFSTVWKKFIAGILAI 121
    +++ +A IG+F+ + P+ GSGIP +EG + GM W+ ++ KF G+ A+
    Sbjct: 81 FLISAFLA--FIGYFLVHRFAPEAAGSGIPEIEGAMDMRPFVRWVRVLPVKFFGGMGALG 138

35  Query: 122 SGLMLGREGGPSIQLGAMTGKGIQYLNASRMEKR-VLIASGAAAGLSAAFNAPIAGLLFV 180
    SG++LGREGP++Q+G G+ I+ + R L+A+GAA GL+AAFNAP+AG++FV
    Sbjct: 139 SGMVLGREGPTVQMGGAVGRMISDIFRVKNEDTRHSLAAGAAGGLAAAFNAPLAGIMFV 198

40  Query: 181 VEEIYHFS-RLVWITALVASLV-ANFVSLNIFGLTPVLALPSELPSLNLFYWFLLMG 238
    +EE+ F L+ + A++ S V AN V I G V+ +P + + L+ +FLL+G
    Sbjct: 199 IEEMRPQFRYTLISVRAVLIISAVAANIVFRVINGQDAVITMP-QYDAPELSTLGLFLLLG 257

45  Query: 239 LFLGILGFIYEWVIL----RFHVIYDYLKFLHLPshlyGILAVIFILPIGYFFPQLLGG 294
    G+ G ++ ++I F + K + L + G + +L Y P+L GG
    Sbjct: 258 ALFGVGFVLFNYLITLAQDLFVKFHRNDRKRYLLTGSMIGGCFGLLLL----YVPELTGG 313

50  Query: 295 GNGLIVSLPRSNLSLMMGLFFLIRFLWSMLSYSGLPGGIFLPILALGSLAG-AFFAVG 353
    G LI ++ +L L F+ R ++L + SG PGGIF P+LALG+L G AF +
    Sbjct: 314 GISLIPTITNGGYGAGILLLLFVGRIFTTLLCFGSGAPGGIFAPMLALGTLFGYAFGLIA 373

55  Query: 354 MQYFGIISHQQISLFFVLGMAGYFGAISKAPLTAMILVTEMVGDQLQMAIGIVTMVSYI 413
    +F ++ + +F + GM F A +AP+T ++LV EM + ++ + I ++ + I
    Sbjct: 374 KMWFPPELNIEP-GMFAIAGMGALFAATVRAPITGILLVIEMTNNYHLILPLIITSLGAVI 432

    Query: 414 VMDLLKGEPIYEAMLAKMTFNP 436
    LL G+PIY +L + N K
    Sbjct: 433 FAQLLGGQPIYSQLLHRTLKNQK 455

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5303> which encodes the amino acid sequence <SEQ ID 5304>. Analysis of this protein sequence reveals the following:

```

60  Possible site: 31
    >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -11.68    Transmembrane    71 - 87 ( 66 - 95)
    INTEGRAL    Likelihood = -9.45     Transmembrane    36 - 52 ( 26 - 56)

```

-1921-

INTEGRAL	Likelihood = -5.63	Transmembrane	346 - 362 (342 - 367)
INTEGRAL	Likelihood = -5.36	Transmembrane	376 - 392 (375 - 393)
INTEGRAL	Likelihood = -5.15	Transmembrane	413 - 429 (410 - 432)
INTEGRAL	Likelihood = -5.10	Transmembrane	321 - 337 (318 - 340)
INTEGRAL	Likelihood = -4.19	Transmembrane	203 - 219 (202 - 220)
INTEGRAL	Likelihood = -4.19	Transmembrane	244 - 260 (242 - 265)
INTEGRAL	Likelihood = -4.19	Transmembrane	284 - 300 (280 - 304)
INTEGRAL	Likelihood = -1.86	Transmembrane	177 - 193 (176 - 194)

----- Final Results -----

bacterial membrane --- Certainty=0.5670(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF96429 GB:AE004383 conserved hypothetical protein [Vibrio cholerae]
Identities = 144/442 (32%), Positives = 236/442 (52%), Gaps = 30/442 (6%)

Query: 18 NEFTFSNKSIIAYVWRGVVVGIIAGVIVSLFRLLIEVTADWVIEWYRYAHINSLLLLPIL 77
N+F +K+ + ++ ++VGI+AG++ + F + + ++ +W + + I S L L +
Sbjct: 21 NQFLSKDKTPFSVLFLSLLVGILAGLVGTYFEQAVHLVSETRTDWLK-SEIGSFLPLWLA 79

Query: 78 SVSLLAVL-FVGFLV--KSDSDIKGSGIPHVEGELKGLMSPDWWSVLWKKFLGGIMAISM 134
+ + A L F+G+ + + + GSGIP +EG + G+ WW VL KF GG+ A+
Sbjct: 80 AFLISAFIAFIGYFLVHRFAPEAAGSGIPEIEGAMDGMRPVRWVRVLPVKFFGGMGALGS 139

Query: 135 GFMLGREGPSIQLGAMSAKGLAKFLKSSRLEKR-VLIASGAAAGLSAAFNAPIAGLLFVV 193
G +LGREGP++Q+G + ++ + + R L+A+GAA GL+AAFNAP+AG++FV+
Sbjct: 140 GMVLGREGPTVQMGGAVGRMISDIFRVKNEDTRHSLLAAGAAGGLAAAFNAPLAGIMFVI 199

Query: 194 EEIYHHFS-RLIWITALVASLV-ANFISLNIIFGLKPVLMSEAMPFLGLNQYWL LLLGL 251
EE+ F LI + A++ S V AN + I G V+ M + L+ L LLLG
Sbjct: 200 EEMRPQFRYTLISVRAVIISAVANIVFRVINGQDAVITMPQ-YDAPELSTLGLFLLGA 258

Query: 252 FLGCLGYLYEIVIL-----NFNKLYVILGSWLHLPDYFYGIIMVFLILPIGYYL 300
G G L+ +I N K Y++ GS + +G++++ Y+
Sbjct: 259 LFGVFGVLFNYLITLAQDLFVKFHRNDRKRYLLTGSMI---GGCFGLLLL-----YV 307

Query: 301 PQLLGGGHGLILSLSNQQLPLMTIFFYFIIRFIVSMFSYSGLEPGGIFLPILTGLAGL 360
P+L GGG LI +++N + F+ R ++ +GSG PGGIF P+L LG L G
Sbjct: 308 PELTGGGISLIPTITNGGYGAGILLLFVGRIFTLLCFGSGAPGGIFAPMLALGTFLFY 367

Query: 361 LFGQIASQLGLLNQSFSLSLFLLGMAGYFAAISKAPLTGMILVTEMVGLKPLMAIAVVT 420
FG IA +F I GM FAA +AP+TG++LV EM + ++ + + +
Sbjct: 368 AFGLIAKMWFPENIEPGMFAIAGMGALFAATVRAPITGILLVIEMTNNYHLILPLIITS 427

Query: 421 FVSYLVMDDLNGQPIYEAMLDK 442
+ + LL GQPIY +L +
Sbjct: 428 LGAVIFAQLLGGQPIYSQLLHR 449

An alignment of the GAS and GBS proteins is shown below.

Identities = 343/510 (67%), Positives = 410/510 (80%)

Query: 1 MENHKNEFQFSLESILGFVWRGIVVGLIAGFVVSIFRLAIEKIFLVVMELYKSAHYQPII 60
MENHKNEF FS +SI+ +VWRG+VVG+IAG +VS+FRL IE V+E Y+ AH ++
Sbjct: 13 MENHKNEFTFSNKSIIAYVWRGVVVGIIAGVIVSLFRLLIEVTADWVIEWYRYAHINSL 72

Query: 61 LLSITVTSTIIAAVIIGFFIKSDPDIKGSGIPHVEGELKGLMSPDWFSIVWKKFIAGILAI 120
LL I S++A + +GF +KSD DIKSGSGIPHVEGELKG++SPDW+S++WKKF+ GI+AI
Sbjct: 73 LLPILSVSLLAVLFVGFVLKSDSDIKGSGIPHVEGELKGLMSPDWWSVLWKKFLGGIMAI 132

Query: 121 SSGMLGREGPSIQLGAMTGKGIAYLNASRMEKRVLIASGAAAGLSAAFNAPIAGLLFV 180
S G MLGREGPSIQLGAM+ KG+A++L +SR+EKRVLIASGAAAGLSAAFNAPIAGLLFV
Sbjct: 133 SMGFMGLGREGPSIQLGAMSAKGLAKFLKSSRLEKRVLIASGAAAGLSAAFNAPIAGLLFV 192

Query: 181 VEETIYHHFSRLVWITALVASLVANFVSLNIFGLTFVLALPSELPSLNLFYWFLLMGLF 240

-1922-

VEEYIHHFSRL+WITALVASLVANF+SLNIFGL PVLA+ +P L LN YW+ LL+GLF
 Sbjct: 193 VEEYIHHFSRLIWITALVASLVANFISLNIFGLKPVLAMSEAMPFLGLNQYWLLLLGLF 252

Query: 241 LGILGFIYEWVILRFHVIYDYLGLKFLPSHLYGILAVIFILPIGYFPQLLGGGGLIV 300
 LG LG++YE VIL F+ +Y LG HLP + YGI+ V ILPIGY PQLLGGG+GLI+
 Sbjct: 253 LGCLGYLYEIVILNFKLYVILGSWLHLDPDYFYGIIMVFLILPIGYLPQLLGGGHGLIL 312

Query: 301 SLPRSNLSLMMGLFLIRFLWSMLSYSSGLPGGIFLPILALGSLAGAFFAVGMQYFGII 360
 SL L LM + +F+IRF+ SM SY SGLPGGIFLPIL LG+LAG F G++
 Sbjct: 313 SLSNQQLPLMTIFFYFIIRFIVSMFSYSGSLPGGIFLPILTLGALAGLLFGQIASQLGLL 372

Query: 361 SHQQISLFFVLGMAGYFGAISKAPLTAMILVTEMVGDLDKQLMAIGIVTMVSYIVMDLLKG 420
 + +SLF++LGMAGYF AISKAPLT MIIVTEMVGDLDK LMAI +VT VSY+VMDLL G
 Sbjct: 373 NQSFLSLFLILGMAGYFAAISKAPLTGMILVTEMVGDLDKPLMAIAVVTFVSYLVMDLLNG 432

Query: 421 EPIYEAMLAQMTFNPDKVMTPTLIELTVSDKISGKYVRDLELPENVLITTQIHHKTSV 480
 +PIYEAML KM ++ PTLIELTV DKI+GKYV++L+LPENVLITTQIHH S V
 Sbjct: 433 QPIYEAMLDKMAKHPTNLVEPTLIELTVGDKIAGKYVKELKLPENVLITTQIHHQKSQV 492

Query: 481 VSGNTILNAGDTIFLVVNESEIKEVREQLM 510
 VSGNT L +G TIFLVVNE++ VRE LM
 Sbjct: 493 VSGNTRLGATIFLVVNEADTGFFVREVM 522

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1707

A DNA sequence (GBSx1811) was identified in *S.agalactiae* <SEQ ID 5305> which encodes the amino acid sequence <SEQ ID 5306>. This protein is predicted to be spermidine/putrescine-binding periplasmic protein precursor (potD-1). Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -9.02 Transmembrane 20 - 36 (14 - 40)

----- Final Results -----
 bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8881> which encodes amino acid sequence <SEQ ID 8882>
 was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 2
 SRCFLG: 0
 McG: Length of UR: 22
 Peak Value of UR: 4.16
 Net Charge of CR: 2
 McG: Discrim Score: 18.94
 GvH: Signal Score (-7.5): -3.29
 Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 Amino Acid Composition: calculated from 1
 ALOM program count: 1 value: -9.02 threshold: 0.0
 INTEGRAL Likelihood = -9.02 Transmembrane 7 - 23 (1 - 27)
 PERIPHERAL Likelihood = 6.05 170
 modified ALOM score: 2.30
 icml HYPID: 7 CFP: 0.461

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1923-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAF94581 GB:AE004221 spermidine/putrescine ABC transporter,
periplasmic spermidine/putrescine-binding protein [Vibrio cholerae]
Identities = 126/327 (38%), Positives = 196/327 (59%), Gaps = 2/327 (0%)

Query: 42 SSSTPNSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSEAMHTKIKQGGITYDIA 101
+++ +L YNW +YI +L+ FTKETGI+V Y T++SNE+M+ K+K G YD+
10 Sbjct: 18 TNAMAKDQELYFYFNWSEYIPSEVLEDFTKETGIKVIYSTYESNESMYAKLKTQAGAGYDLV 77

Query: 102 VPSDYIMDKMIKENLLVKLDHISKIANWDAIGARFKNLSFDPKNKYSIPYFWGTGIVYN- 160
VPS Y + KM KE +L ++DHSK++++ + + N FDP NK+SIPY WG GI N
15 Sbjct: 78 VPSTYFVSKMRKEGMLQEIDHSLKSHFKDLDPNYLNKPFDPGNKFSIPYIWGATGIGINT 137

Query: 161 DQLVKTPPKHWDLLWRPEFRNKIMLVDSAREVIGVGLNSLGYGLNTKNISELKAASKKLD 220
D L K K+W DLW ++ ++ML+D AREV + L+ LGY NT N E+KAA ++L
Sbjct: 138 DMLDKKSLKNWGLDWDKAGQMLMDDAREVFHIALSKLGYSPNTTNPKEIKAAYRELK 197

20 Query: 221 ALTPNVKAIVADEMKGMIQGDAAIGVTFSGEAREMLDGNKHLHYVVPSEGSNLWFDNIV 280
L PNV +D + G+ ++G+ ++G A + + P +G+ W D+I
Sbjct: 198 KLMPNVLVFNSDFPANPYLAGVSLGMLWNGSAYMARQEGAPIQLIWPEKGTIFWMDNIS 257

25 Query: 281 IPKTVKHKREAYAFINFMPEKNAQAQNAEYIGYATPNLKAKALLPADIKNDKAFYPPDKT 340
IP K+ + A+ I+F++ P+NAA+ A IGY TP A LLP + ND + YPP
Sbjct: 258 IPAGAKNIEAAHKMIDFLLRPENAAKIALEIGYPTPVKTAHDLLEKEFANDPSIYPPQSV 317

Query: 341 IDHLEVYNNLGQKWLGIYNDLYLQFKM 367
ID+ E + +G+ + +Y++ + + K+
30 Sbjct: 318 IDNGEWQDEVGEASV-LYDEYFQKLKV 343

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5307> which encodes the amino acid sequence <SEQ ID 5308>. Analysis of this protein sequence reveals the following:

Possible site: 22
35 >>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -8.44 Transmembrane 8 - 24 (1 - 27)

----- Final Results -----
40 bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 >GP:AAC74207 GB:AE000212 spermidine/putrescine periplasmic transport
protein [Escherichia coli]
Identities = 134/342 (39%), Positives = 199/342 (58%), Gaps = 3/342 (0%)

Query: 17 ILTSLSFILQKKSGSGSQSDKLVYINWGDYIDPALLKKFTKETGIEVQYETFDSEAMYT 76
+L + + L + ++ L YNW +Y+ P LL++FTKETGI+V Y T++SNE MY
50 Sbjct: 8 LLAAGALALGMSAAHADDNNTLYFYFNWTEYVPPGLLEQFTKETGIKVIYSTYESNETMYA 67

Query: 77 KIKQ-GGTTYDIAVPDYITDKMIKENLLNKLKSKLVGMDNIGKEFLGKSFDPPQNDYSL 135
K+K YD+ VPS Y +DKM KE ++ K+DKSKL N+ + L K FDP NDYS+
Sbjct: 68 KLKTYKDGAYDLVVPSTYYVDKMRKEGMIQKIDKSKLTNFSNLDPDMLNKPDPNDYSI 127

55 Query: 136 PYFWGTGIVYNDQLVD-KAPMHWDLWRPEYKNSIMLIDGAREMLGVGLTTFGYSVNSK 194
PY WG I N VD K+ W DLW+PEYK S++L D ARE+ + L GYS N+
Sbjct: 128 PYIWGATAIGVNGDAVDPKSVTSWADLWKPEYKGSLLLTDDAREVFQMALRKLGYSGNTT 187

60 Query: 195 NLEQLQAERKLQQLTPNVKAIVADEMKGMIQGDAAIGITFSGEASEMLDSNEHLHYIV 254
+ ++++AA +L++L PNV A +D ++G+ +G+ ++G A + + +
Sbjct: 188 DPKEIEAAYNELKKLMPNVAAPNSDNPANPYMEGEVNLGMIWNGSAFVARQAGTPIDVW 247

Query: 255 PSEGSNLWFDNLVLPKTMKHEKEAYAFNFINRPENAAQNAAYIGYATPNKKAKALLPDE 314

-1924-

P EG W D+L +P K+++ A +NF+ RP+ A Q A IGY TPN A+ LL E
 Sbjct: 248 PKEGGIFWMDSLAIPANAKNKEGALKLINFLLRPDVAKQVAETIGYPTPNLAARKLLSPE 307

Query: 315 IKNDPAFYPTDDIICKLEVYDNLGSRWLGIYNDLYLQFKMYR 356

+ ND YP + IK E +++G+ IY + Y + K R

Sbjct: 308 VANDKTLYPDAETIKNGEWQNDVGAA-SSIYEYYQKLKAGR 348

An alignment of the GAS and GBS proteins is shown below.

Identities = 270/357 (75%), Positives = 306/357 (85%)

Query: 14 MRRVYSFLGGIVLVILILFGLTLYLEKKSSSTPNSDKLVIYNWGDYIDPALLKKFTKETG 73

MR++YSFL G++ VI+IL L+ L+KKS S SDKLVYINWGDYIDPALLKKFTKETG

Sbjct: 1 MRKLYSFLAGVLGVIVILTSLSFILQKSGSGSQSKLVIYNWGDYIDPALLKKFTKETG 60

Query: 74 IEVQYETFDSDNEAMHTKIKQGGTTYDIAVPSDYDIDMIKENLLVKLDHISKIANWDAIGA 133

IEVQYETFDSDNEAM+TKIKQGGTTYDIAVPSDY IDMIKENLL KLD SK+ D IG

Sbjct: 61 IEVQYETFDSDNEAMYTNIKQGGTTYDIAVPSDYDIDMIKENLLNKLKSKLVGMDNIGK 120

Query: 134 RFKNLSFDPKKNYSIPYFWGTVGIVYNDQLVKTPPKHWDLLWRPEFRNKIMLVDSAREVI 193

F SFD+P+N YS+PYFWGTVGIVYNDQLV P HW+DLWRPE++N IML+D ARE++

Sbjct: 121 EFLGKSFDPPQNDYSLPYFWGTVGIVYNDQLVDKAPMHWDLLWRPEYKNSIMLIDGAREML 180

Query: 194 GVGLNSLGYGGLNTKNISELKAASKKLDALTPNVKAIVADEMKGMIQGDAAIGVTFSGEA 253

GVGL + GY +N+KN+ +L+AA +KL LTPNVKAIVADEMKGMIQGDAAIG+TFSGEA

Sbjct: 181 GVGLTTFGYSVNSKNLEQLQAAERKLQQLTPNVKAIVADEMKGMIQGDAAIGITFSGEA 240

Query: 254 REMLDGNKHLHYVVPSEGSNLWFDNIVIPKTVKHKRKEAYAFINFMMEPKNAAQNAEYIGY 313

EMLD N+HLHY+VPSEGSNLWFDN+V+PKT+KH KEAYAF+NF+ P+NAAQNA YIGY

Sbjct: 241 SEMLDNSNEHLHYIVPSEGSNLWFDNVLVLPKTMKHEKEAYAFINFINRPNAAQNAAYIGY 300

Query: 314 ATPNLKAKALLPADIKNDKAFYPPDKTIDHLEVYNNLGQKWLGIYNDLYLQFKMYRK 370

ATPN KAKALLP +IKND AFYP D I LEVY+NLG +WLGIYNDLYLQFKMYRK

Sbjct: 301 ATPNKKAKALLPDEIKNDPAFYPTDDIICKLEVYDNLGSRWLGIYNDLYLQFKMYRK 357

SEQ ID 8882 (GBS135) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 6; MW 40kDa).

GBS135-His was purified as shown in Figure 201, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1708

A DNA sequence (GBSx1812) was identified in *S.agalactiae* <SEQ ID 5309> which encodes the amino acid sequence <SEQ ID 5310>. This protein is predicted to be spermidine/putrescine ABC transporter, permease protein (potC). Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -12.05 Transmembrane 17 - 33 (10 - 37)

INTEGRAL Likelihood = -8.65 Transmembrane 236 - 252 (232 - 259)

INTEGRAL Likelihood = -7.75 Transmembrane 137 - 153 (132 - 158)

INTEGRAL Likelihood = -7.17 Transmembrane 63 - 79 (60 - 92)

INTEGRAL Likelihood = -6.32 Transmembrane 108 - 124 (107 - 136)

----- Final Results -----

bacterial membrane --- Certainty=0.5819(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1925-

A related GBS nucleic acid sequence <SEQ ID 8883> which encodes amino acid sequence <SEQ ID 8884> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1   Crend: 2
SRCFLG: 0
5  McG: Length of UR:   26
    Peak Value of UR:   3.65
    Net Charge of CR:   2
    McG: Discrim Score:   16.58
    GvH: Signal Score (-7.5): -6.17
10  Possible site: 43
    >>> Seems to have an uncleavable N-term signal seq
    Amino Acid Composition: calculated from 1
    ALOM program   count: 4 value: -12.05 threshold: 0.0
    INTEGRAL      Likelihood = -12.05   Transmembrane   9 - 25 ( 2 - 29)
15  INTEGRAL      Likelihood = -7.75   Transmembrane  129 - 145 ( 124 - 150)
    INTEGRAL      Likelihood = -7.17   Transmembrane   55 - 71 ( 52 - 84)
    INTEGRAL      Likelihood = -6.32   Transmembrane  100 - 116 ( 99 - 128)
    PERIPHERAL    Likelihood = 0.53     174
    modified ALOM score: 2.91
20  icml HYPID: 7 CFP: 0.582

    *** Reasoning Step: 3

    ----- Final Results -----
25  bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

30  >GP:AAB91527 GB:AE001165 spermidine/putrescine ABC transporter,
    permease protein (potC) [Borrelia burgdorferi]
    Identities = 97/249 (38%), Positives = 159/249 (62%), Gaps = 3/249 (1%)

    Query: 10  KKFANIYLALVFIIILYIPITIIYLFYSFNKGGMNSFTGFTFSHYGELFQDSRLMLILVQT 69
35  + F NI+L L+ +Y+PII LI YSFN G + GF+ Y E+F S++ + T
    Sbjct: 3   RAFKNIFLFLILSFIIYLPITIIYSFNSGDSGFIWQGFSLKWYKEIFASSQIKSAIFNT 62

    Query: 70  FFLAFLSALLATIIGTFGAIWIYQVRRRH-QTSILSLNNILLVAPDVMIGASFLLVFTVI 128
40  +A +S+L + +IG GA IY+ + +T+LS+N I ++ PD++ G S + ++ I
    Sbjct: 63  ILIAIISLTSVVIIGIIGAYAIYKSENNKCLKTILLSVNKITIINPDIVTGISLMTFYSAI 122

    Query: 129 GLQLGFTSVLLSHVAFSIPIVLMVLPRLKEMNDDMINASYDLGASTWQMLKEVMLPYLS 188
    +QLGF+++L+SH+ FS P VV+++LP+L + ++I+A+ DLGAS Q+ ++ P ++
50  Sbjct: 123 KMQLGFSTMLISHIIFSTPYVIIILPKLYSLPKNIIDAAKDLGASEIQIFFNIIYPEIA 182

    Query: 189 SGIISGFFMAFTYSLDDFAVTFVFTGNGFSTLSVEIYSRARRGISLEINALSTIVF--LF 246
    I +G +AFT S+DDF ++FF TG GF+ LS+ I S +RGI INA+S I+F +
    Sbjct: 183 GSIATGALIAFTLSIDDFLISFFTTGQGFNNLSILINSLTKRGIKPVINAISAILFFFTIL 242

    Query: 247 SILLVIGYY 255
    S+L +I +
50  Sbjct: 243 SLLFIINKF 251

```

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5311> which encodes the amino acid sequence <SEQ ID 5312>. Analysis of this protein sequence reveals the following:

```

    Possible site: 49
    >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL      Likelihood = -8.17   Transmembrane   9 - 25 ( 4 - 29)
60  INTEGRAL      Likelihood = -8.12   Transmembrane  228 - 244 ( 224 - 250)
    INTEGRAL      Likelihood = -7.91   Transmembrane  129 - 145 ( 124 - 150)
    INTEGRAL      Likelihood = -7.06   Transmembrane   62 - 78 ( 54 - 87)
    INTEGRAL      Likelihood = -3.93   Transmembrane  100 - 116 ( 99 - 118)

```

-1926-

----- Final Results -----

bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAB91527 GB:AE001165 spermidine/putrescine ABC transporter,
 permease protein (potC) [Borrelia burgdorferi]
 Identities = 91/249 (36%), Positives = 154/249 (61%), Gaps = 3/249 (1%)

Query: 2 KKFANLYLASVFLLYIPIFYLFYFNFNKGDMNGFTGFTLEHYQTMFEDSRLMTILLQT 61
 + F N++L + +Y+PI LI YSFN G + GF+L+ Y+ +F S++ + + T
 Sbjct: 3 RAFKNIFLFLILSFIYLPILIIYISFNNGSGFIWQGFSLKWKYKEIFASSQIKSAIFNT 62

Query: 62 FVLAFSSALLATIIGIFGAIFIHVRGK-YQNAMLSANNVLMVSPDVMIGASFLILFTSL 120
 ++A S+L + +IGI GA I+ K + +LS N + +++PD++ G S + ++++
 Sbjct: 63 ILIAIISSLTSSVIGIIGAYAIYKSENKKLKTILLSVNKITIINPDIVTGISLMTFYSAI 122

Query: 121 KFQLGMSSVLLSHIAFSIPIVVLMLPRLKEMNQDMVNAAYDLGANYFQMLKEVMLPYFT 180
 K QLG S++L+SHI FS P VV+++LP+L + +++++AA DLGA+ Q+ ++ P
 Sbjct: 123 KMQLGFTMLISHIIFSTPYVVIILPKLYSLPKNIIDAADLGASEIQIFFNIIYPEIA 182

Query: 181 PGIIAGYFMAFTYSLDDFAVTFFLTGNSTVTTLSVEIYSRARGISLDINALSTIVFF--F 238
 I G +AFT S+DDF ++FF TG LS+ I S ++GI INA+S I+FF
 Sbjct: 183 GSIATGALIAFTLSIDDFLISFFTTGQGFNNLSILINSLTKRGIKPVINAISAILFFTIL 242

Query: 239 SILLVIGYY 247
 S+L +I +
 Sbjct: 243 SLLFIINKF 251

An alignment of the GAS and GBS proteins is shown below.

Identities = 196/258 (75%), Positives = 231/258 (88%)

Query: 9 MKKFANIYLALVFIIYIPIIYLFYFNFNKGDMNSFTGFTFSHYGELFQDSRLMLILVQ 68
 MKKFAN+YLA VF++LYIPI YLIFYFNFNKGDMN FTGFT HY +F+DSRLM IL+Q
 Sbjct: 1 MKKFANLYLASVFLLYIPIFYLFYFNFNKGDMNGFTGFTLEHYQTMFEDSRLMTILLQ 60

Query: 69 TFFLAFLSALLATIIGTFGAIWIYQVRRRHQTSILSLNNILLVAPDVMIGASFLLVFTVI 128
 TF LAF SALLATIIG FGAI+I+ VR ++Q ++LS NN+L+V+PDVMIGASFL++FT +
 Sbjct: 61 TFLAFSSALLATIIGIFGAIFIHVRGKYQNAMLSANNVLMVSPDVMIGASFLILFTSL 120

Query: 129 GLQLGFTSVLLSHVAFSIPIVVLMLPRLKEMNDMDINASYDLGASTWQMLKEVMLPYLS 188
 QLG +SVLLSH+AFSIPIVVLMLPRLKEMN DM+NA+YDLGA+ +QMLKEVMLPY +
 Sbjct: 121 KFQLGMSSVLLSHIAFSIPIVVLMLPRLKEMNQDMVNAAYDLGANYFQMLKEVMLPYFT 180

Query: 189 SGIISGFFMAFTYSLDDFAVTFVFTGNSTVTTLSVEIYSRARRGISLEINALSTIVFLFSI 248
 GII+G+FMAFTYSLDDFAVTFV+TGN +TLSVEIYSRAR+GISL+INALSTIVF FSI
 Sbjct: 181 PGIIAGYFMAFTYSLDDFAVTFFLTGNSTVTTLSVEIYSRARGISLDINALSTIVFFFSI 240

Query: 249 LLVIGYYYISKEKGEKNA 266
 LLVIGYYY+S++K EK+A
 Sbjct: 241 LLVIGYYYMSQDKKEKHA 258

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1709

A DNA sequence (GBSx1813) was identified in *S.agalactiae* <SEQ ID 5313> which encodes the amino acid sequence <SEQ ID 5314>. This protein is predicted to be spermidine/putrescine ABC transporter, permease protein (potB). Analysis of this protein sequence reveals the following:

Possible site: 35

-1927-

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -9.55	Transmembrane	250 - 266 (244 - 269)
INTEGRAL	Likelihood = -3.93	Transmembrane	148 - 164 (146 - 166)
INTEGRAL	Likelihood = -3.35	Transmembrane	65 - 81 (64 - 85)
INTEGRAL	Likelihood = -1.97	Transmembrane	96 - 112 (96 - 115)

----- Final Results -----

bacterial membrane	---	Certainty=0.4821(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9853> which encodes amino acid sequence <SEQ ID 9854> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC22990 GB:U32813 spermidine/putrescine ABC transporter,
permease protein (potB) [Haemophilus influenzae Rd]
Identities = 90/255 (35%), Positives = 153/255 (59%), Gaps = 11/255 (4%)

Query: 21 AWLFLFVLAPVALIAWNSFFDINGH-----FTLANYQTFSSGTYLKMSFNSVLYAGIV 74
+WL FVL P L+ SF +G T+ NY F+ Y ++ +NS+ +GI
Sbjct: 18 SWLIFFVLIPNLLVLAVSFLTRDGSNFYAFPITIENTNLFNP-LYAQVVWNSLSMSGIA 76

Query: 75 SFITLLISYPAAYLLTKL--KHKQLWMLVLPTWINLLKAYAFMGIFGQQGGINAFLT 132
+ I LLI YP A++++K+ K++ L L LV+LP W N L++ Y G +G +N L
Sbjct: 77 TIICLLIGYPFAFMMSKIHPKYRPLLLFLVLPFWTNSLIRIYGMKVFLGVKGILNTMLI 136

Query: 133 FIGI--GPKQILFTDFSFLFVAAYIELPFMLLPINFALDDIDQNIYASDDLGANAWQTF 190
+GI P +IL T+ + + Y+ LPFM+LP+++A++ +D L+ A+ DLGAN +Q F
Sbjct: 137 DMGILSAPIRILNTEIAVIIGLVYLLLPFMILPLYSAIEKLDNRLLAARDLGANTFQRF 196

Query: 191 QKVIPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAEQHFLLITQNKMGST 250
+VI PL++ G+ AG V +P++ +F + L+GG +V+ +G I+ FLI++N GS
Sbjct: 197 FRVILPLTMPGIIAGCLLVLLPAMGMFYVADLLGGAKVLLVGNVIKSEFLISRNPWPFGSA 256

Query: 251 IGVILILVMVAIMWL 265
+ + L ++M ++++
Sbjct: 257 VSIGLTVLMALLIFV 271

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5315> which encodes the amino acid sequence <SEQ ID 5316>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -7.38	Transmembrane	19 - 35 (11 - 40)
INTEGRAL	Likelihood = -6.79	Transmembrane	250 - 266 (245 - 268)
INTEGRAL	Likelihood = -4.83	Transmembrane	65 - 81 (63 - 85)
INTEGRAL	Likelihood = -1.97	Transmembrane	96 - 112 (96 - 115)
INTEGRAL	Likelihood = -1.91	Transmembrane	148 - 164 (148 - 165)

----- Final Results -----

bacterial membrane	---	Certainty=0.3951(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

>GP:AAC22990 GB:U32813 spermidine/putrescine ABC transporter,
permease protein (potB) [Haemophilus influenzae Rd]
Identities = 91/262 (34%), Positives = 158/262 (59%), Gaps = 11/262 (4%)

Query: 20 FLWILFFVVPVTLFFYKSFFDIEGR-----VTLANYETFFSSWTYLRMSVNSILYAGI 73
F W++FFV+ P L+ SF +G +T+ NY F+ Y ++ NS+ +GI
Sbjct: 17 FSWLIFFVLIPNLLVLAVSFLTRDGSNFYAFPITIENTNLFNP-LYAQVVWNSLSMSGI 75

-1928-

Query: 74 ITLVTLISYPTALFLTRL--KHKQLWLMLIILPTWVNLLLKAYAFMGIFGQGGINSFL 131
 T++ LLI YP A ++++ K++ L L L++LP W N L++ Y G +G +N+ L
 Sbjct: 76 ATICLLIGYPFAFMMSKIHPKYRPLLLFLVVLFPFWTNSLIRIYGMKVFLGVKGILNTML 135

5 Query: 132 TFMGI--GPQQILFTDFSIFVASYIELPFMMLPIFNALDDIDHNVINASRDLGASEFQA 189
 MGI P +IL T+ + I Y+ LFFM+LP+++A++ +D+ ++ A+RDLGA+ FQ
 Sbjct: 136 IDMGILSAPIRILNTEIAVIIGLVYLLLPFMILPLYSIAIEKLDNRLLLEAARDLGANTFQR 195

10 Query: 190 FSKVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAEQHFLLITQNWGMGS 249
 F +VI PL++ G+ AG V +P++ +F + L+GG +V+ +G I+ FL ++NW GS
 Sbjct: 196 FFRVILPLTMPGIIAGCLLVLLPAMGMFYVADLLGGAKVLLVGNVIKSEFLISRNWPPFGS 255

Query: 250 TIGVVLILTMVAIMWLTKEKSK 271
 + + L + M ++++ +K
 15 Sbjct: 256 AVSIGLTVLMALLIFVYYRANK 277

An alignment of the GAS and GBS proteins is shown below.

Identities = 215/266 (80%), Positives = 239/266 (89%)

20 Query: 4 RRREMKTSSLFSIPYMAWLFLVLAVALIAWNSFFDINGHFTLANYQTFFSSGYLKM 63
 RR MKKTSSLFSIPY W+ FV+APV L+ + SFFDI G TLANY+TFFSS TYL+M
 Sbjct: 4 RRSVMKTSSLFSIPYFLWILFFVAVPTLLFYKSFFDIEGRVTLANYETFFSSWTYLRM 63

25 Query: 64 SFNSVLYAGIVSFITLLISYPAAAYLLTKLKHQLWLMLVILPTWINLLLKAYAFMGIFGQ 123
 S NS+LYAGI++ +TLLISYP A LT+LKHQLWLML+ILPTW+NLLLKAYAFMGIFGQ
 Sbjct: 64 SVNSILYAGIITLVTLISYPTALFLTRLKHQLWLMLIILPTWVNLLLKAYAFMGIFGQ 123

30 Query: 124 QGGINAFITFIGIGPKQILFTDFSFLVAAYIELPFMMLPIFNALDDIDQNLIASDDL 183
 QGGIN+FLTF+GIGP+QILFTDFS+FVA+YIELPFM+LPIFNALDDID N+I AS DLG
 Sbjct: 124 QGGINSFLTTFMGIGPQILFTDFSIFVASYIELPFMMLPIFNALDDIDHNVINASRD 183

35 Query: 184 ANAWQTFQKVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAEQHFLLITQ 243
 A+ +Q F KVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAEQHFLLITQ
 Sbjct: 184 ASEFQAFSKVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAEQHFLLITQ 243

Query: 244 NKGMGSTIGVILILVMVAIMWLTKEK 269
 N GMGSTIGV+LIL MVAIMWLTKE+
 Sbjct: 244 NWGMGSTIGVVLILTMVAIMWLTKEK 269

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1710

A DNA sequence (GBSx1814) was identified in *S.agalactiae* <SEQ ID 5317> which encodes the amino acid sequence <SEQ ID 5318>. This protein is predicted to be spermidine/putrescine ABC transporter, ATP-binding protein (potA). Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.3031(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:AAB91525 GB:AE001165 spermidine/putrescine ABC transporter,
 ATP-binding protein (potA) [Borrelia burgdorferi]
 Identities = 166/345 (48%), Positives = 240/345 (69%), Gaps = 1/345 (0%)

60 Query: 1 MTNPPIIAFKNVSKVFEDSNTVVLKDINFELEEGKFYTLGASGSGKSTILNIIAGLLEAS 60
 M N I+ KN+S ++++ L +IN ++++ +F TLLG SG GK+T++ I+ G L

-1929-

Sbjct: 1 MDNCILEIKNLSHYDNNNGNKTLDNINLKIKKNEFITLLGPGSGCKTTLIKILGGFLSQK 60

Query: 61 TGDIIYLDGKRINDVPTNKRQDVHTVFQNYALFPHMTVFENVAFFLKLKMDKKEIQKRVQE 120
G+IY K I+ NKR+++TVFQNYALFPHM VF+N++F L++KK K I+++V+

5 Sbjct: 61 NGEIYFFSKEISKTSPNKREINTVFQNYALFPHMNVFDNISFGLRMKKTTPKDIKEKVKVT 120

Query: 121 TLKMRVLEGFEEKRAIQKLSGGQQRVAIARAIINQPKVLLDEPLSALDLKLRTEMQYEL 180
+L ++ + R I +LSGGQ+QRVAIARA++ +PK++LLDEPLSALDLK+R EMQ EL

10 Sbjct: 121 SLSLIGMPKYAYRNINELSGGQQRVAIARAMVMEPKLLLLDEPLSALDLKMRQEMQKEL 180

Query: 181 RELQORLGITFVFVTHDQEEALAMSDWIFVMNEGEIVQSGTPVDIYDEPINHFVATFIGE 240
+++Q++LGITF++VTHDQEEAL MSD I VMNEG I+Q GTP +IY+EP FVA FIGE

Sbjct: 181 KKIQRQLGITFIYVTHDQEEALTMSDRIVVMNEGIIILQIGTPEEIYNEPKTKFVADFGE 240

15 Query: 241 SNILSGKMIEDYLVEFNGKRFEAVDGGMRPNESVQVIRPEDLQITLPDEGKLQVKVDTQ 300
SNI G ++ +V G FE +D G E+V +VIRPED+++ +G L + +

Sbjct: 241 SNIFDGTYYKELVVSLLGHEFECLDKGFEAEAEVDLIRPEDVKLLPKGKGHLSGTITSA 300

Query: 301 LFRGVHYEIIAYDDLGNEMMIHSTRKAIEGEVIGLDFTPEDIHIM 345
+F+GVHYE+ N W++ STR GE + + P+DIH+M

20 Sbjct: 301 IFQGVHYEMTLEIQKTN-WIVQSTRLTQVGEVDIFLEPDDIHVM 344

There is also homology to SEQ ID 1292

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
25 vaccines or diagnostics.

Example 1711

A DNA sequence (GBSx1815) was identified in *S.agalactiae* <SEQ ID 5319> which encodes the amino acid sequence <SEQ ID 5320>. Analysis of this protein sequence reveals the following:

Possible site: 53
30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4990(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06283 GB:AP001515 UDP-N-acetylenolpyruvoglucosamine
reductase [Bacillus halodurans]
40 Identities = 119/286 (41%), Positives = 166/286 (57%), Gaps = 1/286 (0%)

Query: 13 DIRFDEPLKKYTYTKVGGPADYLAFFPRNRLSRIKVFANSQNIPWMVLGNASNIIVRDG 72
++R +E L +T K+GGPAD P + L +K W V+G SNI+V D

45 Sbjct: 15 EVRVNESLAHHTTWKIGGPADVFPNDIEGLKNTMKLIQETGCKWRVIGRGSNIIIVSDK 74

Query: 73 GIRGFVIMFDK-LSTVTVNGYVIEAEAGANLIETTRIARYHSLTGFEFACGIPGSVGGAV 131
G+RG I DK L + VNG I AG +++ + L G EFA GIPGSVGGAV

Sbjct: 75 GLRGVTIKLDKGLDHLEVNAGESITVGAGFPVVKLATVISRQGLAGLEFAAGIPGSVGGAV 134

50 Query: 132 FMNAGAYGGEIAHILLSAQVLTPQGELEKTIERNMQFGYRHSVIQESGDIVISAKFALKP 191
FMNAGA+G +I+ IL A VL P G L+ + M F YR S++Q++ I + A F+L

Sbjct: 135 FMNAGAHGSDISQILTKAHVLPDGTLRWLNEEMAFSYRTSLQKNDGICVEAIFSLTR 194

55 Query: 192 GDHLMITQEMDRLTYLRELKQPLEYPSGVSFVKRPPGHFAGQLISEAHLKGQRIGGVEVS 251
GD I +++ + R QP +P+CGSVF+ P +AGQLI +A LKG +IGG ++S

Sbjct: 195 GDKEDIKKLQKNKDYRRDTQPNHPTCGSVFRNPLPEYAGQLIEKAGLKG YQIGGAQIS 254

Query: 252 QKHAGFMVNIAEGSAQDYENLIEHVINTVESTSGVHLEPEVRIIGE 297
HA F+VN + A D LI HV +T++ +++E EV +IGE

60 Sbjct: 255 TMHANFIVNTGDAKAADVLALIHVKTDTIQKQYQNMETEVELIGE 300

-1930-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5321> which encodes the amino acid sequence <SEQ ID 5322>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4557(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 229/292 (78%), Positives = 267/292 (91%)

```

Query: 8  ELEGLDIRFDEPLKKYTYTKVGGPADYLAFFPRNRLELSRIVKFANSQNIPWMVLGNASNI 67
          EL G+DIR +EPLK YTYTKVGGPAD+LAFPRN ELSRIV +AN +N+PW+VLGNASN+
Sbjct: 4  ELHGIDIRENEPLKHYYTYTKVGGPADFLAFPRNHYELSRIVAYANKENMPWLVLGNASNL 63

Query: 68  IVRDGGIRGFVIMFDKLSVTVNGYVIEAEAGANLIETTRIARYHSLTGFEFACGIPGSV 127
          IVRDGGIRGFVIMFDKL+ V +NGY +EAEAGANLIETT+IA++HSLTGFEFACGIPGS+
Sbjct: 64  IVRDGGIRGFVIMFDKLNVAHLNGYTLAEAGANLIETTKIAKFHSLTGFEFACGIPGSI 123

Query: 128  GGAVFMNAGAYGGEIAHILLSAQVLTPOGELKTIEARNMQFGYRHSVIQESGDIVISAKF 187
          GGAVFMNAGAYGGEI+HI LSA+VLTP GE+KTI AR+M FGYSHS IQE+GDIVISAKF
Sbjct: 124  GGAVFMNAGAYGGEISHIFLSAKVLTPSGEIKTISARDMAFGYRHSIQETGDIVISAKF 183

Query: 188  ALKPGDHLMITQEMDRLTYLRELKQPLEYPCSGSVFKRPPGHFAGQLISEAHLKGQRIGG 247
          ALKPG++ I+QEM+RL +LR+LKQPLE+PCSGSVFKRPPGHFAGQLI EA+LKG RIGG
Sbjct: 184  ALKPGNYDTISQEMNRLNHLRQLKQPLEFPSCGSVFKRPPGHFAGQLIMEANLKGHRIGG 243

Query: 248  VEVSQKHAGFMVNIAEGSAQDYENLIEHVINTVESTSGVHLEPEVRIIGESL 299
          VEVS+KH GFM+N+A+G+A+DYE+LI +VI TVE+ SGV LEPEVRIIGE+L
Sbjct: 244  VEVSEKHTGFMINVADGTAKDYEDLIAYVIETVENHSGVRLEPEVRIIGENL 295

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1712

A DNA sequence (GBSx1816) was identified in *S.agalactiae* <SEQ ID 5323> which encodes the amino acid sequence <SEQ ID 5324>. This protein is predicted to be 2-amino-4-hydroxy-6-hydroxymethyldihydropterin pyrophosphokinase/dihyd. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1122(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA03814 GB:AP001507

2-amino-4-hydroxy-6-hydroxymethyldihydropteridin e
pyrophosphokinase [Bacillus halodurans]

Identities = 64/146 (43%), Positives = 94/146 (63%)

```

Query: 5  YLSLGSNIGDRETFLKQALFSIDHLQKTKVAQISAIYETAAGNTNQEDFFNICQVETD 64
          Y++LGSNIGDR FL++A+ + K V S+IYET G T+Q F N+ +V T
Sbjct: 6  YIALGSNIGDRSRFLEEAIQQLAEHDKVTVTCCSSIYETDPVGYTDQSPFLNMVVEVSTS 65

```

-1931-

Query: 65 LAPFELLDYQCQIEKCLKRVREHGWGPRITIDILLFGNQVINQEDLVVPHPYMTKRAFV 124
 L +LL+ Q+IE+ R RH WGPRT+D+DILL+ + E+L++PHP M +RAFV
 Sbjct: 66 LPVEQLLEVTQKIERYCGRERHWRGPRITDLDILLYDQENREMNLIIPHPMWERAFV 125

5 Query: 125 LVPLLEIAPQLSLPNGSKLEDYLEKL 150
 L+PL+E+ P + P+G +E + +L
 Sbjct: 126 LIPLMELNPSIVAPSGKTIEQVVREL 151

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5325> which encodes the amino acid
 10 sequence <SEQ ID 5326>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.0479(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

20 Identities = 85/156 (54%), Positives = 111/156 (70%), Gaps = 1/156 (0%)

Query: 1 MTTVYLSLGSNIGDRETFLKQALFSDHLOKTKVAQISAIYETAANGNTNQEDFFNICCQ 60
 MT VYLSLG+N+GDR +L++AL ++ L +T++ S+IYET AWG T Q DF N+ CQ
 Sbjct: 1 MTIVYLSLGTNMGDRAAYLQKALEALADLPQTRLLAQSSIIYETTAWGKTGQADFLNMACQ 60

25 Query: 61 VETDLAPFELLDYQCQIEKCLKRVREHGWGPRITIDILLFGNQVINQEDLVVPHPYMTK 120
 ++T L + L Q IE+ L RVRHE WG RTIDIDILLFG +V + ++L VPHPYMT+
 Sbjct: 61 LDTQLTAADFLKETQAIEQSLGRVRHEKWSRTIDIDILLFGEEVYDTKELKVPHPYMTE 120

30 Query: 121 RAFVLVPLLEIAPQLSLPNGSK-LEDYLEKLNLGEV 155
 RAFVL+PLLE+ P L LP K L DYL L+ ++
 Sbjct: 121 RAFVLIPLELQPDLLPPNHKFLRDYLAALDQSDI 156

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 35 vaccines or diagnostics.

Example 1713

A DNA sequence (GBSx1817) was identified in *S.agalactiae* <SEQ ID 5327> which encodes the amino
 acid sequence <SEQ ID 5328>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2826(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5329> which encodes the amino acid
 sequence <SEQ ID 5330>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.3547(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-1932-

Identities = 75/119 (63%), Positives = 92/119 (77%)

Query: 1 MDKIYLNKCRFYGYHGAFSEEQTLGQVFQVDAVLSLDLAKASQTDDLIDTVHYGEVFD CI 60
 MDKI L CRFYGYHGAF EEQTLGQ+F VD LS+DL AS +D L DTVHYG VFD +
 5 Sbjct: 1 MDKIVLEGCRCFYGYHGAFKEEQTLGQIFLVDLELSVDLQAASLSDQLTDTVHYGMVFD SV 60

Query: 61 KNHVENEQYQLIEKLAGVIVEDIFLQFHPVQAITLKITKDNPPINGHYESVGIELERRR 119
 + VE E++ LIE+LAG I E +F +F P++AI + I K+NPPI GHY++VGIELER+R
 10 Sbjct: 61 RQLVEGEKFILIERLAGAICEQLFNEFPPIEAIKVAIKKENPPIAGHYKAVGIELERQR 119

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1714

A DNA sequence (GBSx1818) was identified in *S.agalactiae* <SEQ ID 5331> which encodes the amino acid sequence <SEQ ID 5332>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5333> which encodes the amino acid sequence <SEQ ID 5334>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 181/267 (67%), Positives = 224/267 (83%), Gaps = 1/267 (0%)

Query: 1 MKIGQYDITGKACIMGILNVTDPDSFSDGGSYTTIDSALNQVGEMLEQGVAVIDIGGESTR 60
 MKIG++ I G A IMGILNVTDPDSFSDGGSYTT+ AL+ V +M+ G I+D+GGESTR
 40 Sbjct: 1 MKIGKFVIEGNAAIMGILNVTDPDSFSDGGSYTTVQKALDHVEQMIADGAKIIDVGGESTR 60

Query: 61 PGAVFVTAEEIEIKRVVPMIKAIREVYPDLLLSIDTYKTEVAQAALDAGVHILNDVWVSGLY 120
 PG FV+A +EI RVVP+IKAI+E Y D+L+SIDTYKTE A+AAL+AG ILNDVW+GLY
 Sbjct: 61 PGCQFVSATDEIDRVVPVIAIKENY-DILISIDTYKTETARAALDAGADILNDVWAGLY 119

Query: 121 DGKMLSAAERNVPIILMHNQEEAVYQDIKKEVCEFLLEAERALEAGVSKDNIWIDPGF 180
 DG+M +LAAE + PIILMHNQ+E VYQ++ ++VC+FL RA+ AL+AGV K+NIW+DPGF
 Sbjct: 120 DGQMFALAAEYDAPILMHNQDEEVYQEVTDVCDFLGNRAQAALDAGVPKNNIWDVDPGF 179

Query: 181 GFAKTEEQNLELLKGLEQVCDLGPVFLFGISRKRITVNYLLGGNREVTERDMGTAALSAYA 240
 GFAK+ +QN ELLKGL++VC LGYPVFLFGISRKR V+ LLGGN + ERD TAALSA+A
 50 Sbjct: 180 GFAKSVQONTPELLKGLDRVCQLGYPVFLFGISRKRVDALLGGNTKAKERDGAALSAAYA 239

Query: 241 IAKGCQIVRVHNVENKDIVTVISQLV 267
 + KGCQIVRVH+V+ N+DIV V+SQL+

Sbjct: 240 LGKGCQIVRVHVDVKNQDIVAVLSQLM 266

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1933-

Example 1715

A DNA sequence (GBSx1819) was identified in *S.agalactiae* <SEQ ID 5335> which encodes the amino acid sequence <SEQ ID 5336>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2429(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5337> which encodes the amino acid sequence <SEQ ID 5338>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1590(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 151/184 (82%), Positives = 166/184 (90%)

Query: 3 NQEKMEKAIYQFLEALGENPNREGLKDTPKRVAKMYIEMFSGLNQDPKEQFTAVFSENHE 62

N+EK E AIYQFLEA+GENPNREGL DTPKRVAKMY EMF GL +DPKE+FTAVF E HE

Sbjct: 16 NKEKABAAIYQFLEAIGENPNREGLDTPKRVAKMYAEMFLGLGKDPKEEFTAVFKEQHE 75

Query: 63 EVVIVKDIPIFYSCHEHLVPFYGKAHIAYLPNDGRVTGLSKLARAVEVASKRPQLQERLT 122

+VVIVKDI FYS+CEHHLVPFYGKAHIAYLP+DGRVTGLSKLARAVEVASKRPQLQERLT

Sbjct: 76 DVVIVKDISFYSCHEHLVPFYGKAHIAYLPDGRVTGLSKLARAVEVASKRPQLQERLT 135

Query: 123 AQVAQALDALAPKGIFVMIEAEHMCMTMRGIKKPGSKTITTTVARGLYKDDRYERQEILS 182

+Q+A AL +AL PKG VM+EAEHMCMTMRGIKKPGSKTITTT ARGLYK+ R ERQE++S

Sbjct: 136 SQIADALVEALNPKGTLVMVEAEHMCMTMRGIKKPGSKTITTTTARGLYKESRAERQEVIS 195

Query: 183 LIQK 186

L+ K

Sbjct: 196 LMTK 199

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1716

A DNA sequence (GBSx1820) was identified in *S.agalactiae* <SEQ ID 5339> which encodes the amino acid sequence <SEQ ID 5340>. This protein is predicted to be folylpolyglutamate synthase (folC). Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2836(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9855> which encodes amino acid sequence <SEQ ID 9856> was also identified.

-1934-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14768 GB:Z99118 folyl-polyglutamate synthetase [Bacillus subtilis]
Identities = 154/426 (36%), Positives = 245/426 (57%), Gaps = 17/426 (3%)

5   Query: 3   YQEALEWIHISKLAFGIKPGLERMRWMLEQLGNPQNLSAIHVVGTTNGKGGSTTSYLQHI 62
      YQ+A WIH +L FG+KPGL RM+ ++ +LG+P+ + A HV GTNGKGGST +++++ +
      Sbjct: 5   YQDARSWIHGRLKFGVKPGLGRMKQLMARLGHPEKKIRAFHVAGTNGKGGSTVAFIRSMQL 64

10  Query: 63  NSGYQVGTFTSPYIVDFRERISIDQMIPESDFIKLVETVRPVVERLHLETNLEPATEFE 122
      +GY VGTFTSPYI+ F ERIS++G I + ++ LV ++P VE L +T TEF
      Sbjct: 65  EAGYTVGTFTSPYIITFNERISVNGIPISEDEWTALVNQMKPHVEALD-QTEYGPTEFE 123

15  Query: 123  VITVLMFYFFGNSCPVDIVIEAGMGYYDSTNMFKALAVTCPSIGLDHQEVLGRITYVDI 182
      ++T F YF VD VI E G+GG +DSTN+ + L SIG DH +LG T +I
      Sbjct: 124  IMTACAFLYFAEFHKVDFVIFETGLGGRFDSTNVVEPLLTVITSIGHDHMMNILGNTIEEI 183

20  Query: 183  AEQKVGVLKKGVPFVYANDRQDVVEVFQIKAKETHSQTYRLHNDFYIKEEE-----NYFN 237
      A +K G++K+G+P V A + + +V + +A+ + LH+ I EE F+
      Sbjct: 184  AGEKAGIIKEGIPIVTAVTQPEALQVIRHEAERHAAPFQSLHDACVIFNEEALPAGEQFS 243

25  Query: 238  YIGPQANIDHIQLQMPGHHQVSNASIAI-TTSLLLRDKYPKLTQLTIKDGLEMTKWVGRT 296
      + + + I+ + G HQ NA+++I L ++ ++ + ++ GL W GR
      Sbjct: 244  FKTEEKCYEDIRTSLIGTHQRQNAALSILAAEWLNKENIAHISDEALRSGLVKAAPGRL 303

30  Query: 297  ELI--FPNVMIDGAHNHNSVDALVQVIK-KYQQKNVHILFAAINTKPIESMLESLSIA- 352
      EL+ P V +DGAHN E V+ L + +K ++ + ++F+A+ KP ++M++ L +IA
      Sbjct: 304  ELVQEHPPVYLDGAHNEEGVEKLAETMKQRFANSRISVVSALKDKPYQNMIRLETIAH 363

35  Query: 353  PVSVTSFDYPK-SINLDKYPKAYTRVSDWKWLHDI-----NLTSKDFYVITGSLYFIS 406
      + SFD+P+ S+ D Y + W + D+ + + +ITGSLYFIS
      Sbjct: 364  AIHFASFDFPRASLAKDLYDASEISNKSWSDEPDDEVKFIKSKGSNEIVLITGSLYFIS 423

Query: 407  QVRQEL 412
      +R+ L
Sbjct: 424  DIRKRL 429
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5341> which encodes the amino acid sequence <SEQ ID 5342>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
40  >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL Likelihood = -1.28 Transmembrane 12 - 28 ( 12 - 28)

----- Final Results -----
45  bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 230/411 (55%), Positives = 295/411 (70%), Gaps = 1/411 (0%)

50  Query: 1   MTYQEALEWIHISKLAFGIKPGLERMRWMLEQLGNPQNLSAIHVVGTTNGKGGSTTSYLQHI 60
      MTY+E LEWIIH L FGIKPGL+RM W+L QLGNPQ N+ +H+VGTNGKGGST ++LQHI
      Sbjct: 34  MTYEETLEWIIHDHLVFGIKPGLKRLWVLGQLGNPQKNVKGHVHIVGTNGKGGSTVNHLQHI 93

55  Query: 61  FTNSGYQVGTFTSPYIVDFRERISIDQMIPESDFIKLVETVRPVVERLHLETNLEPATE 120
      FT +GY+VGTFTSPYI+DF+ERISI+G+MI E D + +RP+ ERL ET+ TE
      Sbjct: 94  FTTAGYEVGTFTSPYIMDFKERISINGRMISEKDLVIAANRIRPLTERLVQETDFGEVTE 153

60  Query: 121  FEVITVLMFYFFGNSCPVDIVIEAGMGYYDSTNMFKALAVTCPSIGLDHQEVLGRITYV 180
      FEVIT++MF YFG+ PVDI IIEAG+GG YDSTN+F+A+ V CPSIGLDHQ +LG TY
      Sbjct: 154  FEVITLIMFLYFGDMHPVDIAIIEAGLGGLYDSTNVFQAMVVVCP SIGLDHQAILGETYA 213

Query: 181  DIAEQKVGVLKKGVPFVYANDRQDVVEVFQIKAKETHSQTYRLHNDFYIKEEENFYNYIG 240
      +IA QK GVL+ G V+A + EVF KA++ + + F + E + + +
```


-1935-

Sbjct: 214 NIAAQKAGVLEGGETLVFAVENPSAREVFLTKAEQVGASIWEEQEQFQMAENASGYRFTS 273

Query: 241 PQANIDHIQLQMPGHHQVSNASIAITTSLLLRDKYPKLTTLQTIKDGLEMTKWVGRTELIF 300
P I I + MPGHHQVSNAAI T L L+D+YP+LT I++GL + W+GRTEL+

5 Sbjct: 274 PLGVISDIHIAMPGHHQVSNAAALAIMTCLTLQDRYPRLTPDHIREGLANSLWLGRTELLA 333

Query: 301 PNVIMIDGAHNNEVDALVQVIK-KYQQKNVHILFAAINTKPIESMLESLSSIAPVSVTSF 359
PN+MIDGAHNNEV ALV V+K Y K +HILF AI+TKPI ML +L I + VTSF

10 Sbjct: 334 PNLIMIDGAHNNEVAAALVAVLKNNYNDKKLHILFGAIDTKPIADMLVALEQIGDLQVTSF 393

Query: 360 DYPKSINLDKYPKAYTRVSDWKWLHDINLTSDKDFYVITGSLYFISQVRQ 410
YP + L+KYP+ + RV+D+K +L DF+VITGSLYFIS++RQ

Sbjct: 394 HYPNAYPLEKYPFRVADFKDFLALRKHAKADDFVITGSLYFISEIRQ 444

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1717

- 20 A DNA sequence (GBSx1821) was identified in *S.agalactiae* <SEQ ID 5343> which encodes the amino acid sequence <SEQ ID 5344>. This protein is predicted to be *rarD*. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have a cleavable N-term signal seq.

25 INTEGRAL Likelihood = -12.31 Transmembrane 130 - 146 (125 - 151)
INTEGRAL Likelihood = -10.24 Transmembrane 269 - 285 (262 - 291)
INTEGRAL Likelihood = -7.75 Transmembrane 212 - 228 (207 - 233)
INTEGRAL Likelihood = -5.52 Transmembrane 80 - 96 (75 - 99)
INTEGRAL Likelihood = -4.14 Transmembrane 106 - 122 (104 - 125)
INTEGRAL Likelihood = -3.50 Transmembrane 182 - 198 (180 - 204)
INTEGRAL Likelihood = -2.44 Transmembrane 40 - 56 (39 - 57)
30 INTEGRAL Likelihood = -0.96 Transmembrane 153 - 169 (152 - 169)
INTEGRAL Likelihood = -0.32 Transmembrane 251 - 267 (250 - 267)

----- Final Results -----

35 bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:BAB07585 GB:AP001520 unknown conserved protein [Bacillus halodurans]
Identities = 109/288 (37%), Positives = 185/288 (63%), Gaps = 6/288 (2%)

Query: 7 GIILGLSAYVLWGLLSLYWKLSSGIEAYSTFAYRIIFTVLTLMIYMLVSGRKTVYVKDLK 66
G+I +SAY++WG L LYWKL+ + A A+RI++++ M+I + V + ++++

45 Sbjct: 8 GVIAAISAYLIWGLFLPLYWKLVDVPAEMLAHRIVWSLGFMVILLAVMKKNRQVMREIL 67

Query: 67 GLVNNKKSFWTMFVASILISINWLVIYIFAVTHGHATEASLGYYMMPISILLSVLVLRH 126
+ NKK+ + + VA+ILIS+NW ++I+AV+ EASLGYY+ P+I++LL+++ LRE

Sbjct: 68 DTLANKKTAFGITVAAILISMNWFIFIVAVSSDKVIEASLGYYINPLINVLLAIVFLRES 127

50 Query: 127 LARVVSILAILIIMGVILVYQTHGFPLISLTIALSFGFYGLLKKSISLSSDFSMLVESS 186
L++ + L+A GV + G FP ++ LA+SFG YGL+KK +SLS+ S+ +E+

Sbjct: 128 LSKWEVASFLAAAGVLNITLHYGSFPWVAFALAISFGVYGLIKKVVSLSAWASLTITL 187

55 Query: 187 FIAPFALIYIVFF-----AKDFLTIDYNIQLVLLSLSGIITAVPLLLFAEAIKRAPLNII 241
+ PFAL+++++ A F ++ + L+ SG TA+PLLLFA KR ++I

Sbjct: 188 IMTPFALLFLLYIPLSGGASAFSLNH-LSTAWLIASGAATALPILLFATGAKRISFSLI 246

Query: 242 GFIQYINPTIQLLLALFIFKETIVSGEVIGFIFIWLAILVFSIGQVHT 289
GF+QY+ PTI L+L +F+F+E + + F+ IW +++F+I + T

60 Sbjct: 247 GFLQYLAPTIMLMGVFLFQEPFSRVQFVSFLLIWTGLIIFTISRRT 294

-1936-

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8885> and protein <SEQ ID 8886> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 3
5  McG: Discrim Score:      5.30
   GvH: Signal Score (-7.5): -1.64
      Possible site: 38
   >>> Seems to have a cleavable N-term signal seq.
ALOM program count: 9 value: -12.31 threshold: 0.0
10  INTEGRAL Likelihood ==-12.31 Transmembrane 130 - 146 ( 125 - 151)
    INTEGRAL Likelihood ==-10.24 Transmembrane 269 - 285 ( 262 - 291)
    INTEGRAL Likelihood = -7.75 Transmembrane 212 - 228 ( 207 - 233)
    INTEGRAL Likelihood = -5.52 Transmembrane 80 - 96 ( 75 - 99)
15  INTEGRAL Likelihood = -4.14 Transmembrane 106 - 122 ( 104 - 125)
    INTEGRAL Likelihood = -3.50 Transmembrane 182 - 198 ( 180 - 204)
    INTEGRAL Likelihood = -2.44 Transmembrane 40 - 56 ( 39 - 57)
    INTEGRAL Likelihood = -0.96 Transmembrane 153 - 169 ( 152 - 169)
    INTEGRAL Likelihood = -0.32 Transmembrane 251 - 267 ( 250 - 267)
20  PERIPHERAL Likelihood = 7.96 229
    modified ALOM score: 2.96

*** Reasoning Step: 3

----- Final Results -----
25  bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

ORF02052(319 - 1152 of 1485)
GP|9654601|gb|AAF93371.1||AE004110(13 - 289 of 302) rarD protein {Vibrio cholerae}
%Match = 20.4
%Identity = 37.7 %Similarity = 66.3
Matches = 104 Mismatches = 89 Conservative Sub.s = 79

KDIVNLW*RNLK**NKSALKMVRMLLICLEQDRR*WFCVRKKKNQLSQS*VNYV*VDRFKCLILSEKE*ELRKDNLGII
MFMTPDQQDAKKGIL
10

LGLSAYVLWGLLSLYKWLSSGIEAYSTFAFYRII--FTVLTMLIYMLVSGRKTIVYLKDCLKLVNNKKSFWTMVFASILISI
LALSAYTMMGIAPIFYFKALGAUSAILEILSHRVVSFVLLAVLIHLGRWRWSVV-----GVVHTPRKFLLLVTALLVG
30 40 50 60 70 80

NWLVIYIFAVTHGHATEASLGYIMMPIISILLSVLRLREHARVLSAILIAIMGVIGLYQTGFHFLISLTIALSFGFYG
NWLFIFIWSINANHMDLASLGYIYNPLLNVLGMLFI GERLRKLQWFAVALAAIGVGIQLVVFVGSVPIVAIALATSFGFYG
100 110 120 130 140 150 160

LLKKSISLSSDFSMLEVSSFALIPIYIFFAKDFLDY--NILQL-VLLSLSGIITAVPLLFAEAIKRAPLNIIGFI
LLRKKIQVDQAOTGLFLETLMFLPAAAIYLIWLADTPTSMAINTWQLNLLLVLCAGVVTTLPLLCFTGAAARKLKSLTGFF
180 190 200 210 220 230 240

QYINPTIQLLALLAFIFKETIVSGEVIGFIFIWLAILLVFSIGQVHTMLKKGK*DDLRSARMDS**ISFWY*TRFGTYEMD
QYIGPSLMLFLAVLVYGAEFTSDKAITFAFIWSALVIFSVDGLKAGAARRAR
260 270 280 290 300

-1937-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1718

- 5 A DNA sequence (GBSx1822) was identified in *S.agalactiae* <SEQ ID 5345> which encodes the amino acid sequence <SEQ ID 5346>. Analysis of this protein sequence reveals the following:

Possible site: 34
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.5200(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1719

- 20 A DNA sequence (GBSx1823) was identified in *S.agalactiae* <SEQ ID 5347> which encodes the amino acid sequence <SEQ ID 5348>. Analysis of this protein sequence reveals the following:

Possible site: 20
>>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0881(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 30 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC44297 GB:U41735 homoserine kinase homolog [Streptococcus pneumoniae]
Identities = 188/289 (65%), Positives = 232/289 (80%), Gaps = 1/289 (0%)

- 35 Query: 1 MRIIVPATSANIGPGFDSIGVALSKYLIIEVLEESTEWLVEHNLVN-IPKDHTNLLIQTA 59
M+IIIVPATSANIGPGFDS+GVA++KYL IEV EE EWL+EH + IP D NLL+ A
Sbjct: 1 MKIIIVPATSANIGPGFDSVGVAVTKYLQIEVSEERDEWLVIEHQIGKWIPHDERNLLLTIA 60
- 40 Query: 60 LHVKSIDLAPHRLKMFSDIPLARGLGSSSSVIVAGIELANQLGNLALSQKEKLEIATRLEG 119
L + DL P RLKM SD+PLARGLGSSSSVIVAGIELANQLG L LS EKL++AT++EG
Sbjct: 61 LQIVPDLQPRRLKMTSDVPLARGLGSSSSVIVAGIELANQLGQLNLSDEKQLATKIEG 120
- 45 Query: 120 HPDNVAPAIIFGDLVISSIVKNDIKSLEVMFPDSSFIAFIPNYELKTSRNVLPQKLSYE 179
HPDNVAPAI+G+LVI+S V+ + ++ FP+ F+A+IPNYEL+T DSR+VLP+KLSY+
Sbjct: 121 HPDNVAPAIYGNLVIASSVEGQVSAIVADFPECDFLAYIPNYELRTRDSRSVLPKKLSYK 180
- 50 Query: 180 DAVASSSVANVMVASLLKGDIVTAGWALERDLFHERYRQPLVKEFEVIKQISTQNGAYAT 239
+AVA+SS+ANV VA+LL GD+VTAG AIE DLFHERYRQ LV+EF +IKQ++ +NGAYAT
Sbjct: 181 EAVAASSIANVAVAALLAGDMVTAGQAIEGDLFHERYRQDLVREFAMIKQVTKENGAYAT 240
- Query: 240 YLSGAGPTVMVLCSEKEQAIVTELSKLCGQIQVLNIERKGVRVEKR 288
YLSGAGPTVMVL S +K I EL K G++ L ++ +GVRVE +
Sbjct: 241 YLSGAGPTVMVLASHDKMPTTIKAELEKQPFKGKLDLRVDTQGVREAK 289

No corresponding DNA sequence was identified in *S.pyogenes*.

-1938-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1720

A DNA sequence (GBSx1824) was identified in *S.agalactiae* <SEQ ID 5349> which encodes the amino acid sequence <SEQ ID 5350>. This protein is predicted to be homoserine dehydrogenase (hom). Analysis of this protein sequence reveals the following:

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9857> which encodes amino acid sequence <SEQ ID 9858> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA65713 GB:X96988 hom [Lactococcus lactis]
Identities = 221/432 (51%), Positives = 307/432 (70%), Gaps = 11/432 (2%)

Query: 15  MTKIALLGFGTVAKGIPYLLKENQHKLSSLEGEDIVIDKVLVRDNESRQRFINQGFTYN 74
           M + IA+LGFGTV G+P LL EN+ KL + E+IVI KVL+RDN++ ++ +QGF Y+
Sbjct: 1   MAVNIAILGFGTVGTGLPTLLSENKEKLAKILDEEIVISKVLMRDNKAIEKARSQGFNYD 60

Query: 75  FVTEINTILQDSQIDIVVELMGGIEPAKTYLSQALGFGKHIVTANKDLIALHGKELMDLA 134
           FV ++ IL DS+I IVVELMG IEPAKTY++QA+ GK++VTANKDL+A+HG EL LA
Sbjct: 61  FVLNLDLADSEISIVVELMGRIEPAKTYITQAIEAGKNVVTANKDLLAVHGVELRSLA 120

Query: 135 DARGLALFYEGAVAGGIPILRTLSSHFSASDKMTRLLGILNGTSNFMFLTKMFEEGWSYEQA 194
           +AL+YE AVAGGIPILRTL++SF+SDK+T LLGILNGTSNFM+TKM EEGW+Y+++
Sbjct: 121 QKHHVALYYEAAVAGGIPILRTLANSFSSDKITHLLGILNGTSNFM+TKMSEEGWYDES 180

Query: 195 LKKAQELGYAESDPTNDVEGIDTAYKATILSQFGFGMPIDFDDVNYKGISSIRSEDVEVA 254
           L KAQELGYAESDPTNDV+GID +YK ILS+F FGM + DD+ G+ SI+ DVE+A
Sbjct: 181 LAKAQELGYAESDPTNDVDGIDASYKLAILEFAFGMTLAPDDIAKSGLSIQKTDVEIA 240

Query: 255 QEMGFAIKLVADLRETPTGISVDVSPTLISQKHPLAAVNHVMNAVFIESIGIGQSLFYGP 314
           Q+ G+ +KL ++ E +GI +VSPT + + HPLA+VN VMNAVFIES GIG S+FYG
Sbjct: 241 QQFGYVLKLTGEINEVDSGIFAEVSPFPLPKSHPLASVNGVMNAVFIESEGIGDSVFYGA 300

Query: 315 GAGQNPTATSVLADIIDISRSIRSIQIKIPMNTYHPCRLSMQSDIFNEYYLAI SLRAE 374
           GAGQ PTATSVLADI+ I + ++ K N Y L+ DI N+YY ++ E
Sbjct: 301 GAGQKPTATSVLADIVRIVKRVKDGITIGKSFNEYARSTSLANPHDIENKYFYSV-----E 355

Query: 375 DSDTLGR-----YFEQENIGLKNVIEKALGDKQOEIYVLTDEVSQEKITQFIEEFPEESG 428
           D+ G+ F EN+ + V+++ K+ + +++ +++++ ++ +
Sbjct: 356 TPDSTGQLLLLVELFTSENVSFQVLQKQNGKRAVVVIISHKINRVQLSAIQDKLNQEK 415

Query: 429 VIQLINVFVKVIG 440
           +L+N FKV+G
Sbjct: 416 DFKLLNRFKVLG 427
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1939-

Example 1721

A DNA sequence (GBSx1825) was identified in *S.agalactiae* <SEQ ID 5351> which encodes the amino acid sequence <SEQ ID 5352>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4548(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1722

A DNA sequence (GBSx1826) was identified in *S.agalactiae* <SEQ ID 5353> which encodes the amino acid sequence <SEQ ID 5354>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -12.79 Transmembrane 20 - 36 (14 - 41)

----- Final Results -----

bacterial membrane --- Certainty=0.6116(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15906 GB:Z99123 similar to hypothetical proteins [Bacillus subtilis]
Identities = 105/272 (38%), Positives = 149/272 (54%), Gaps = 20/272 (7%)

Query: 25 FLIALIGIFLFFNNRSKQEIKT-----KTNASSHRKIVTSIKKKK-----WIKQKTPVK 74
FL I L+G L + QE K K ++KK+ WIK + P K

Sbjct: 5 FLSIFLLGSCALAAACDAQEANAQPMKAEQKKPEKKAVQVQKKEDDTSAWIKTEKPAK 64

Query: 75 IPILMYHAVHVM DPSEAA SANLIVAPDIFESHKRLKKEGYFLAPNEAYRALNENALPE 134
+PILMYH++ ++ +L V FE+H+K L GY L P EA L ++ P

Sbjct: 65 LPILMYHSI-----SSGNSLRVPKKEFEAHMKWLHDNGYQTLTPKEASLMLTQDKKPS 117

Query: 135 KKVIWITFDDGNADFYTKAYPILKKYKVKATNNIITGFVQEGRESNLNVQQMLEMKQNGM 194
+K + ITFDDG D Y AYP+LKKY +KAT +I + G + +L +QM EM Q+G+

Sbjct: 118 EKCVLITFDDGYTDNYQDAYPVLKKYGMKATIFMIGKSI--GHKHHLTTEQMKEMAQHGHI 175

Query: 195 SFQGHTVTHPNLSLLTPQLTQEMTL SKQFLDQKLSQDTLAIAYPSGRYNPTTLDIASQY 254
S + HT+ H L+ LTP+ Q EM SK+ D Q T I+YP GRYN TL A +

Sbjct: 176 SIESHTIDHLELNLGLTPQQQSEMA DSKKLFDMFHHQQTIIISYPVGRYNEETLKAAEKT 235

Query: 255 -YKLGLTTNEGVA TKDNGLLSLNRIRILPTTS 285

Y++G+TT G A++D G+ +L+R+R+ P S

Sbjct: 236 GYQMGVTTTEPGAASRDQGMALHRVRVSPGMS 267

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5355> which encodes the amino acid sequence <SEQ ID 5356>. Analysis of this protein sequence reveals the following:

Possible site: 24

-1940-

>>> May be a lipoprotein

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:CAB15906 GB:Z99123 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 97/240 (40%), Positives = 140/240 (57%), Gaps = 9/240 (3%)

Query: 71 KKTHFDSSKSQKKAHSLTWTKQETPVKIPILMYHAIHVMSPEETANANLIVNPDLDQ 130
 KK + + QKK W K E P K+PILMYH+I ++ +L V F+
 Sbjct: 37 KKPEKKAQVQKKEDDTSAWIKTEKPAKLPIILMYHSI-----SSGNSLRVPKKEFEAH 89

15 Query: 131 LQKMKDEGYFLSPPEVYRALSNNELPAKKVVLTFDDSMIDFYNVAYPILKKYDAKATN 190
 ++ + D GY L+P+E L+ ++ P++K V +TFDD D Y AYP+LKKY KAT
 Sbjct: 90 MKWLHDNGYQTLTPKEASMLTQDKPSEKCVLITFDDGYTDNYQDAYPVLKKYGMKATI 149

20 Query: 191 NVITGLTEMGSAANLTLKQMKEMKQVGMFSQDHTVNHDPLEQASPDVQTTMKDSKDYLD 250
 +I +G +LT +QMKEM Q G+S + HT++H +L +P Q +EM DSK D
 Sbjct: 150 FMIG--KSIGHKHLTEEQMKEMAQHGISTESHTIDHLELNLTPQQQSEMADSKKLPD 207

25 Query: 251 KQLNQNTIAIAYPSGRYNDTTLQIAARLNYKLGVTTNEGIASAANGLLSLNRIRILPNMS 310
 +Q T I+YP GRYN+ TL+ A + Y++GVIT G AS G+ +L+R+R+ P MS
 Sbjct: 208 NMFHQQTIIISYPVGRYNEETLKAARKTGVMGVTTPEGAASRDQGMALHRVRVSPGMS 267

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 153/265 (57%), Positives = 199/265 (74%), Gaps = 4/265 (1%)

Query: 33 IFLFFNNRSKQEIATK---TNASSHRKIVTSIKKKKWKIKQKTPVKIPILMYHAVHVMDS 89
 I LF + ++ ++ TK T+ S + + K W KQ+TPVKIPILMYHA+HVM P
 Sbjct: 54 ISLFFHHKKTAKKETTKLKKTHFDSSKSQKKAHSLTWTKQETPVKIPILMYHAIHVMSPE 113

35 Query: 90 EAASANLIVAPDIFESHKRLKKEGYFLAPNEAYRALNENALPEKKVIWITFDDGNADF 149
 E A+ANLIV PD+F+ ++++K EGYFL+P E YRAL+ N LP KKV+W+TFDD DF
 Sbjct: 114 ETANANLIVNPDLDQQLQKMKDEGYFLSPPEVYRALSNNELPAKKVVLTFDDSMIDF 173

40 Query: 150 YTKAYPILKKYKVKATNNIITGFVQEGRESNLNVQMLEMKQNGMSFQGHTVTHPNLSLL 209
 Y AYPILKKY KATNN+ITG + G +NL ++QM EMKQ GMSFQ HTV HP+L
 Sbjct: 174 YNVAYPILKKYDAKATNNVITGLTEMGSAANLTLKQMKEMKQVGMFSQDHTVNHDPLEQA 233

45 Query: 210 TPELQTQEMTLKQFLDQKLSQDTLAIAYPSGRYNPTTLDIASQY-YKLGLTTNEGVA 268
 +P++QT EM SK +LD++L+Q+T+AIAYPSGRYN TTL IA++ YKLG+TTNEG+A+
 Sbjct: 234 SPDVQTTMKDSKDYLDKQLNQNTIAIAYPSGRYNDTTLQIAARLNYKLGVTTNEGIASA 293

50 Query: 269 DNGLLSLNRIILPTTSDDDLIKTI 293
 NGLLSLNRIILP S ++L++T+
 Sbjct: 294 ANGLLSLNRIILPNMSPENLLQTM 318

SEQ ID 5354 (GBS287d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 3 & 4; MW 57kDa) and in Figure 185 (lane 2; MW 57kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 6; MW 32kDa) and in Figure 181 (lane 5; MW 32kDa).

55 Purified GBS287d-GST is shown in Figure 243, lanes 10-11; purified GBS287d-His is shown in Figure 234, lanes 7-8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1941-

Example 1723

A DNA sequence (GBSx1828) was identified in *S.agalactiae* <SEQ ID 5357> which encodes the amino acid sequence <SEQ ID 5358>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 21
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
10   bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1724

A DNA sequence (GBSx1829) was identified in *S.agalactiae* <SEQ ID 5359> which encodes the amino acid sequence <SEQ ID 5360>. Analysis of this protein sequence reveals the following:

```

20   Possible site: 40
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3352 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
25   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1725

A DNA sequence (GBSx1830) was identified in *S.agalactiae* <SEQ ID 5361> which encodes the amino acid sequence <SEQ ID 5362>. This protein is predicted to be glycine betaine transporter BetL (opuD). Analysis of this protein sequence reveals the following:

```

35   Possible site: 61
    >>> Seems to have an uncleavable N-term signal seq

    INTEGRAL    Likelihood = -12.68    Transmembrane  439 - 455 ( 435 - 491)
    INTEGRAL    Likelihood = -12.10    Transmembrane  256 - 272 ( 249 - 281)
    INTEGRAL    Likelihood = -11.30    Transmembrane  464 - 480 ( 456 - 491)
40   INTEGRAL    Likelihood = -10.83    Transmembrane   49 - 65 ( 44 - 74)
    INTEGRAL    Likelihood = -10.40    Transmembrane   11 - 27 ( 5 - 34)
    INTEGRAL    Likelihood = -9.98     Transmembrane  396 - 412 ( 390 - 419)
    INTEGRAL    Likelihood = -9.29     Transmembrane  224 - 240 ( 220 - 247)
    INTEGRAL    Likelihood = -7.11     Transmembrane  347 - 363 ( 341 - 366)
45   INTEGRAL    Likelihood = -2.87     Transmembrane  143 - 159 ( 143 - 159)
    INTEGRAL    Likelihood = -2.60     Transmembrane  192 - 208 ( 191 - 208)
    INTEGRAL    Likelihood = -1.44     Transmembrane   86 - 102 ( 86 - 105)

    ----- Final Results -----
50   bacterial membrane --- Certainty=0.6074 (Affirmative) < succ>

```

-1942-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAD30266 GB:AF102174 glycine betaine transporter BetL [Listeria
 monocytogenes]
 Identities = 277/503 (55%), Positives = 365/503 (72%), Gaps = 1/503 (0%)

10 Query: 4 KHITPVFTGSLIVSLILVLLGIIVPRGFQSWTQILREQVSTNFGWLYLLLVTSILALCVF 63
 K +T VF GS + L+ VL G +P F+++T +++ ++NFGW YL++V I+ C+F
 Sbjct: 2 KKLITNVFWGSGFLVLLAVLFGAFLPEQFETFTNHIQKFLTSNFGWYLLIVVAIIIIIFCLF 61

15 Query: 64 FIMSPLGQIRLGQPHSRPEYSTVSWIAMMFSAGMGIGLVFYGAAEPLSHFAISTPGAPKE 123
 ++SP+G IRLG+P P YS SW AM+FSAGMGIGLVF+GAAEPLSH+A+ PG
 Sbjct: 62 LVLSPIGSIRLGKPGEEPYSNKSFWAMLFSAAGMGIGLVFWGAAEPLSHYAVQAPGGEVG 121

20 Query: 124 SQTALADAFRFTFFHWHGIHAWAVYALVALALAYFGFRKQEKYLLSVTLKPLFGDKTDGWL 183
 +Q A+ DA R++FFHWGI AW++YA+VALALAYF FRK L+S TL P+ G G +
 Sbjct: 122 TQAMKDALRYSFFHWGISAWSIYAIVALALAYFKFRKNAPGLISATLYPILGKHAKGPI 181

25 Query: 184 GKIVDITTVVATVIGVATTGFGAAQINGGLSFLGVPNNAFVQIVIIILITTALFVMSAL 243
 G+++DI V ATVIGVATTG GA QINGGL++L GVPNN VQ II+I T LF++SA+
 Sbjct: 182 GQLIDIIAVFATVIGVATTGLGAQQINGGLTYLFGVPNNFTVQFTIIVIVTILFMLSAM 241

30 Query: 244 SGLGKGVKILSNLNLILAVALLALVIVLGPTVRIFDTLTESLGSYLQNFSGMSFRAAAFD 303
 SGL KG+++LSN+N+ +A LL L ++LGPT+ I + T S G YLQN MSF+ A
 Sbjct: 242 SGLDKGIQLLSNVNIYVAGVLLVLTLLIGPTLFIMNFTNSFGDYLQNIQMSFQTAPDA 301

35 Query: 304 NTKRSWIDNWTIFYWAWWISWSPFVGFIARISKGRSIREFLTIVVLLIPTLLSFVWFAAF 363
 R WID+WTIFYWAWW+SWSPFVG+FIARIS+GR+IR+FL V+++P L+S WFA F
 Sbjct: 302 PDARKWIDSWTIFYWAWWLSWSPFVGFIARISGRRTIRQFLLGIVLPLVSVFVFAVF 361

40 Query: 364 GTLSTQVQQLG-TNLTKFATEEVLFEATFNHYTLGWLLSIIAILIFSFFITSADSATYVL 422
 G + V+Q G + L+ ATE+VLF FN + G +LSI+A+ILI FFITSADSAT+VL
 Sbjct: 362 GGSATFVEQHNSGLSSLAETEQLFGVFNEFPGMMLISIVAMILIAVFFITSADSATFVL 421

45 Query: 423 AMLTEDGNLNPKNRTKVIWGLVLAVIAIVLLLSGGLLALQNVLIIVALPFSFVMILMMLA 482
 M T G+LNP N KV WGL+ A IA VLL +GGL ALQN II A PFS V+ILM+++
 Sbjct: 422 GMQTTGGSLLNPPNSVKVTWGLLQAGIASVLLYAGGLTALQNASIIAAPPFSIVIIILMIVS 481

50 Query: 483 LLVELPFHEKKEMGLSISPDYRPR 505
 L V L E++++GL + P + R
 Sbjct: 482 LRVSLTREQEKLGLYVRPKKSQR 504

45 No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8887> and protein <SEQ ID 8888> were also identified. Analysis of this protein sequence reveals the following:

50 Lipop: Possible site: -1 Crend: 4
 McG: Discrim Score: 15.28
 GvH: Signal Score (-7.5): -4.24
 Possible site: 61

>>> Seems to have an uncleavable N-term signal seq

55 ALOM program count: 11 value: -12.68 threshold: 0.0

INTEGRAL	Likelihood = -12.68	Transmembrane	439 - 455 (435 - 491)
INTEGRAL	Likelihood = -12.10	Transmembrane	256 - 272 (249 - 281)
INTEGRAL	Likelihood = -11.30	Transmembrane	464 - 480 (456 - 491)
INTEGRAL	Likelihood = -10.83	Transmembrane	49 - 65 (44 - 74)
INTEGRAL	Likelihood = -10.40	Transmembrane	11 - 27 (5 - 34)
INTEGRAL	Likelihood = -9.98	Transmembrane	396 - 412 (390 - 419)
60 INTEGRAL	Likelihood = -9.29	Transmembrane	224 - 240 (220 - 247)
INTEGRAL	Likelihood = -7.11	Transmembrane	347 - 363 (341 - 366)
INTEGRAL	Likelihood = -2.87	Transmembrane	143 - 159 (143 - 159)
INTEGRAL	Likelihood = -2.60	Transmembrane	192 - 208 (191 - 208)
INTEGRAL	Likelihood = -1.44	Transmembrane	86 - 102 (86 - 105)

PERIPHERAL Likelihood = 3.50 319
modified ALOM score: 3.04

5

10

```

bacterial membrane --- Certainty=0.6074(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

ORF02057(310 - 1821 of 2145)
 GP|4835822|gb|AAD30266.1|AF102174_1|AF102174(2 - 506 of 507) glycine betaine transporter
 BetL {Listeria monocytogenes} PIR|T48645|T48645 glycine betaine transport protein betL
 [validated] - Listeria monocytogenes
 %Match = 38.7
 %Identity = 54.9 %Similarity = 74.7
 Matches = 277 Mismatches = 127 Conservative Sub.s = 100

20

54 84 114 144 174 204 234 264
 IGGHHYRNYRLEVLKIDMVVS*ANLDMPLSTNIWYLHOIVINH*VKHKNOIMLFGSEFLRROGEVLIQVVKMRGVEF

25

294 324 354 384 414 444 474 504
KVCYTIILV*EEILSKKHITPVFTGSLIVSLILVLGLIIVPRGFQSWTQILREQVSTNFGWLYLLLVTSILALCVFFIMSP
| : | || | : | : || : | : | : :: : : : || | | : : | : : : : ||
MKKLTVNFVSGSGLFVLLAVLFGAFLPEQFETFTNIHQKFLTNSFGWYYLIVVAIIIFCLFLVLSP

10 20 30 40 50 60

30

534 .564 594 624 654 684 714 744
 LGQIRLGQPHSRPEYSTVSWIAMMFSAAGMIGIGLVFYGAABPLSHFAISTPGAPKESQTALADAFRFTFFHWGIHAWAVVA
 :| |||:| || || |:|||||:|||||:|: || |: |:|:||||| |:|
 IGSIRLGKPGEEP GYSNKSWFAMLFSAAGMIGIGLVFWGAABPLSHYAVQAPGGEVGTQAAMKDALRYSFFHWGISAWSIYA
 80 90 100 110 120 130 140

35

774 804 834 864 894 924 954 984
I V A L A L A Y F G F R K Q E K Y L L S V T L K P L F G D K T D G W L G K I V D I T T V V A T V I G V A T T L G F G A A Q I N G G L S F L L G V P N N A F V Q I
:
| | | | | | | | | | : | : | : | : | | | | | | | | | | : | : | : | : | : | | |
I V A L A L A Y F K F R K N A P G L I S A T L Y P I L G K H A K G P I G O L I D I I A V F A T V I G V A T T L G L G A Q I N G G L T Y L F G V P N N F T V Q F

160 170 180 190 200 210 220

40

[illegible]

45

1254 1284 1314 1344 1374 1404 1431 1461
WIDNWTIFYWAWWISWSPPFVGFIARISKGRSIREFLTVLLIPTLLSFVWFFAAGTGLSTQVQQLGTN-LTKFATEEVLV
|||::||| |||||::||| |||||::||| |::| |::| |::| |::| :|| |::| :|| |::| :|| |::|
WIDSWTIFYWAWWLSWSPPFGIFIAIRISGRTRIQLLGVIVLPALVSVFVFAVFGGSAIFVEQHGNSSGLSSLAETQVLF

 320 330 340 350 360 370 380

50

1491 1521 1551 1581 1611 1641 1671 1701
ATFNHYTLGWLLSIIAILILPSFFITTSADSATYVLAMLTEDGNLNPKNRKTQVIWGLVLAIVIAIALLSCGILLALQNVLII
|| : | : ||| : ||| ||||| ||||| ||||| ||||| : ||||| : ||||| : ||||| : ||||| : |||||
GVFNEFP GGMMLISIVAMILIAVF FFI TSADSATFVLGMQT TGGSLNPPNSVKVTWG LLQA GISAV LLYAG GLTAL QNASII

400 410 420 430 440 450 460

55

1731 1761 1791 1821 1851 1881 1911 1941
 VALPFSFVMILMMLALLLVLPHEKKEMGLSISPDYRPRKNKPFKSYYE*KEARLLFIG*SS*SDHHR**LVRYEFD*EK
 | : ||| | : ||| :: : | | | : :: || : | : |
 AAFPFSIVIIIMIVLSLVSFLTREQELGLVVRPKKSQRSQL
 480 490 500

65

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1944-

Example 1726

A DNA sequence (GBSx1831) was identified in *S.agalactiae* <SEQ ID 5363> which encodes the amino acid sequence <SEQ ID 5364>. This protein is predicted to be succinic semialdehyde dehydrogenase (gabD-1). Analysis of this protein sequence reveals the following:

```

5   Possible site: 43
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2733(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9859> which encodes amino acid sequence <SEQ ID 9860> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:AAD19405 GB:AF102543 succinic semialdehyde dehydrogenase
      [Zymomonas mobilis]
    Identities = 229/455 (50%), Positives = 305/455 (66%), Gaps = 5/455 (1%)

20   Query: 10 MAYKTIYPYTNEVLHEFDNISDSLEQSLDIAHALYKTWRKEDNVEERQNLHKVADLLR 69
      MAY+++ P T E + ++ + SD ++ S+D A ++K + + ER LHK A++ R
      Sbjct: 1 MAYESVNPATGETVKKYPDFSDKQVKDSVDRAATVFKNDWSQRTIAERSKVLHKAEIFR 60

25   Query: 70 KDRDKYAEVMTKDMGKLFTEAQGEVDLCADIADYADNGQKFLKVPVPLESPNGEAYYLKQ 129
      D DKYA+++T DMGK EA+GEV+L ADI DYVA NG+KFL P +E G A
      Sbjct: 61 SDVDKYAKLLTIDMGKKIAEARGEVNLSADILDYAKNGEKFLAPQKVEEKPG-AVVKAF 119

      Query: 130 AVGVLLAVEPWNFFPYQIMRVFAPNFIVGNTMLLKHASICPASAQAFEDLVREAGAPEG 189
      +G+LLA+EPWNFP+YQ+ R+ P I GN +L+KH+S P SA AFE ++ EAGAP+G
30   Sbjct: 120 PLGLLLAIEPWNFFPYQLARIAGPYLIAGNALLVKHSSSVPSAHAFAVLEEAGAPKGI 179

      Query: 190 FKNIFASYDQVSNLISDPRVAGVCLTGSEGGASIAAEAGKNLKKSSMELGGNDAFLILD 249
      + N+ AS DQVS +I DPRV GV +TGS GA +AA+AGK KKS MELGG+DAF++LD
35   Sbjct: 180 YTNLDASPDQVSQIIEDPRVRGVTVTGSASVGAEAAKAGKMWKKSVMELGGSDAFIVLD 239

      Query: 250 DADFD--LLSKTIFFARLYNAGQVCTSSKRFFIVMADKYDE-FVMNVVETFKSAKWGDPM 306
      D D L+ K + RL+NAGQV ++KRFI++ K E F + + F++ K GDPMD
      Sbjct: 240 GVDIDDKLIDKAAY-GRLFNAGQVFCAAKRFIIVGQKRAELFTEKLKQRFALKIGDPM 298

40   Query: 307 SETTLAPLSSAGAKDDVLKQIKLAVDHGAEEVVFNDTIDHPGNFVMPVLTNITKANPIY 366
      T L PLSS GA+D V+KQ++ AV +GA++V G I+ G F+ +LT+I + NP Y
      Sbjct: 299 ESTDLGPLSSVGARDQVVKQVEKAVQNGAKLVCGGKAIEGKGAFMKAGILTDIKRENPA 358

      Query: 367 NQEIFGPVASIYKVDTEEBEIALANDSSYGLGSTVFSSDPEHAKKVAAQIETGMTFINSG 426
      +E FGP+A IY V E EAI LANDS YGLG VF+ D E +KVA QIETGM IN
45   Sbjct: 359 FEEFFGPVIAQIYAVKDEAEAEIELANDSPYGLGGAVFAPDVEQGRKVAEQIETGMVAINKP 418

      Query: 427 WTSLELPFPGGIKNSGYGRELSQLGFDAFVNEHLV 461
      + PELPFGG+K+SGYGRELS G F+N L+
50   Sbjct: 419 LWTAPELPFGGVKHSGYGRELSHFGIQEFINWKLI 453

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5365> which encodes the amino acid sequence <SEQ ID 5366>. Analysis of this protein sequence reveals the following:

```

55   Possible site: 27
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2887(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
60   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-1945-

An alignment of the GAS and GBS proteins is shown below.

Identities = 335/457 (73%), Positives = 397/457 (86%)

```

5   Query: 9   IMAYKTIYPYTNEVLHEFDNISDSLEQSLDIAHALYKTRKEDNVEERQNLHKVADLL 68
      +MAY+TIYPYTNEVLH FDN++D L L+ AH LYK WRKED++EER+ QLH+VA++L
      Sbjct: 1   VMAYQTIYPYTNEVLHTFDNMTDQGLADVLERAHLLYKKWRKEDHLEERKAQLHQVANIL 60

10  Query: 69   RKDRDKYAEVMTKDMGKLFTEAQGEVDLCADIADYYADNGQKFLKVPVPLESPNGEAYYLK 128
      R+DRDKYAE+MTKDMGKLFTEAQGEV+LCADIADYYAD +FL PLE+ +G+AYYLK
      Sbjct: 61   RRDRDKYAEIMTKDMGKLFTEAQGEVNLCAADIADYYADKADEFLMSTPLETDSGQAYYLK 120

      Query: 129  QAVGVLLAVEPWNFPFYQIMRVFAPNFIVGNTMLLKHASICPASAQAFEDLVREAGAPEG 188
      Q+ GV+LAVEPWNFP+YQIMRVFAPNFIVGN M+LKHASICP SAQ+FE+LV EAGA G
15  Sbjct: 121  QSTGVILAVEPWNFPYYQIMRVFAPNFIVGNPMVLKHASICPRSAQSFEELVLEAGAEAG 180

      Query: 189  AFKNIFASYDQVSNLISDPRVAGVCLTGSEGGASIAAEAGKNLKKSSMELGGNDAFLIL 248
      + N+F SYDQVS +I+D RV GVCLTGSEGGASIA EAGKNLKK+++ELGG+DAF+IL
      Sbjct: 181  SITNLFISYDQVSQVIADKRVGVCLTGSEGGASIAEEAGKNLKKTTLELGGDDAFIIL 240

20  Query: 249  DDADFDLLSKTIFFARLYNAGQVCTSSKRFIVMADKYDEFVNMVVFETFKSAKWGDPMDSE 308
      DDAD+D L K ++F+RLYNAGQVCTSSKRFIV+ YD F ++ + FK+AKWGDPMDE
      Sbjct: 241  DDADWDQLEKVLVFSRLYNAGQVCTSSKRFIVLDKDYDRFKELLTKVFKTAKWGDPMDE 300

25  Query: 309  TTLAPLSSAGAKDDVLKQIKLAVDHGAEEVVGNDTIDHPGNFVMPTVLTNITKANPIYNQ 368
      TTLAPLSSA AK DVL QIKLA+DHGAE+V+G + IDHPG+FVMPT++ +TK NPIY Q
      Sbjct: 301  TTLAPLSSAQAKADVLDQIKLALDHGAELVYGGEAIDHPGHFVMPTTIAGLTKDNPIYYQ 360

      Query: 369  EIFGPVASIYKVDTEEEAIALANDSSYGLGSTVFSSDPEHAKKVAQIETGMTFINSWGT 428
      EIFGPV IYKV +EEEAI +ANDS+YGLG T+FSS+ EHAK VAA+IETGM+FINSWGT
30  Sbjct: 361  EIFGPVGEIYKVSSEEEAIEVANDSNYGLGGTIFSSNQEHA KAVAAKIETGMSFINSWGT 420

      Query: 429  SLPELPFGGIKNSGYGRELSQLGFDAFVNEHLVFTPN 465
      SLPELPFGGIK+SGYGRELS+LGF +FVNEHL++ PN
35  Sbjct: 421  SLPELPFGGIKHSYGRELSSELGFTSFVNEHLIYIPN 457

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1727

40 A DNA sequence (GBSx1832) was identified in *S.agalactiae* <SEQ ID 5367> which encodes the amino acid sequence <SEQ ID 5368>. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have a cleavable N-term signal seq.

```

45  ----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

50 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1946-

Example 1728

A DNA sequence (GBSx1833) was identified in *S.agalactiae* <SEQ ID 5369> which encodes the amino acid sequence <SEQ ID 5370>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -7.91	Transmembrane	94 - 110 (86 - 115)
INTEGRAL	Likelihood = -7.75	Transmembrane	154 - 170 (150 - 176)
INTEGRAL	Likelihood = -7.11	Transmembrane	316 - 332 (312 - 339)
INTEGRAL	Likelihood = -6.16	Transmembrane	258 - 274 (253 - 278)
INTEGRAL	Likelihood = -2.71	Transmembrane	218 - 234 (217 - 234)
INTEGRAL	Likelihood = -1.49	Transmembrane	286 - 302 (283 - 302)
INTEGRAL	Likelihood = -0.96	Transmembrane	73 - 89 (73 - 89)
INTEGRAL	Likelihood = -0.27	Transmembrane	121 - 137 (121 - 137)

----- Final Results -----

bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9861> which encodes amino acid sequence <SEQ ID 9862> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC75219 GB:AE000305 orf, hypothetical protein [Escherichia coli K12]
Identities = 102/331 (30%), Positives = 172/331 (51%), Gaps = 26/331 (7%)

Query: 12 IPGLILCFIIA-IPSWLLGLYLPLIGAPVF-----AILIGIIVGSFYQNR--QLFNKGIA 63
IPGL L +I + W G +P + F AIL+G+++G+ + + G+
Sbjct: 17 IPGLALSAVITGVALW--GGSIPAVAGAGFSALTLLAILLGMVLGNTIYPHIWKS CDGGVL 74

Query: 64 FTSKYILQTAVVLLGFGNLQMVKVGISSLPPIIMTISISLIIAYVL-QKLFKLDKTIA 122
F +Y+L+ ++L GF L Q+ VGIS + I ++T+S + ++A L QK+F LDK +
Sbjct: 75 FAKQYLLRLGIILYGFRLTFSQIADVGISGIIIDVLTLSSTFLLACFLGQKVFGLDKHTS 134

Query: 123 TLIGVGSSICGSAIAATAPVINAKDDEVAQAISVIFLNFILALIFPTLGNFIG--LSD 180
LIG GSSICG +A+ AT PV+ A+ +V A++ + +F +A ++P + + S
Sbjct: 135 WLIGAGSSICGAAAVLATEPVVKAASKVTAVATVVIFGTVAIFLYPAIYPLMSQWFSP 194

Query: 181 HGFALFAGTAVNDTSSVTAT--ATAWDAINHSNTLGGATIVKLTRTLAIIPITIVLSIYH 238
F ++ G+ V++ + V A A + DA N A I K+ R + + P I+L+
Sbjct: 195 ETFGIYIGSTVHEVAQVVAAGHAISPDAEN-----AAVSKMLRVMM LAPFLILLAA-R 247

Query: 239 MKQTQKEQSVSVTKI-FPKFVLYFILASLLTTIVASLGFSRLRIFEPLKVL SKFFIVMAMG 297
+KQ S +KI P F + FI+ ++ + + L L F + MAM
Sbjct: 248 VKQLSGANSGEKSKITIPWFAIFIVVAIFNSFHL---LPQSVVNMLVTLDTFLLAMAMA 304

Query: 298 AIGINTNVSKLIKTTGGKSILLGAACWLGI 328
A+G+ T+VS L K G K +L+ + +I+
Sbjct: 305 ALGLTTHVSALKKAGAKPLLMALVLF AWLIV 335

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5371> which encodes the amino acid sequence <SEQ ID 5372>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.29	Transmembrane	30 - 46 (22 - 50)
INTEGRAL	Likelihood = -8.12	Transmembrane	314 - 330 (311 - 338)
INTEGRAL	Likelihood = -6.05	Transmembrane	8 - 24 (7 - 29)
INTEGRAL	Likelihood = -6.00	Transmembrane	150 - 166 (146 - 172)
INTEGRAL	Likelihood = -5.57	Transmembrane	257 - 273 (252 - 277)
INTEGRAL	Likelihood = -3.50	Transmembrane	91 - 107 (87 - 108)
INTEGRAL	Likelihood = -2.60	Transmembrane	69 - 85 (68 - 87)
INTEGRAL	Likelihood = -2.55	Transmembrane	289 - 305 (289 - 305)

-1947-

----- Final Results -----

5 bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC75219 GB:AE000305 orf, hypothetical protein [Escherichia coli]
 10 Identities = 100/329 (30%), Positives = 173/329 (52%), Gaps = 21/329 (6%)

Query: 8 LPGLLLCLLLALPAWCLGRLPFIIGAP---VFAILLGMLLA-LFYEHRDKTKEG-ISFT 61
 +PGL L ++ A G + + GA AILLGM+L Y H K+ +G + F
 15 Sbjct: 17 IPGLALSAVITGVALWGGSSIPAVAGAGFSALTLLAILLGMVLGNTIYPHIWKSCDGGVLFA 76

Query: 62 SKYILQTAHVLLGFGLNLTQVMAVGMQSLPIIISTITATALLVAYGL-QKWLRLDVNTATL 120
 +Y+L+ ++L GF L +Q+ VG+ + I + T+++ L+A L QK LD +T+ L
 20 Sbjct: 77 KQYLLRLGIILYGFRLTFSQIADVGISGIIIDVLTLSSTFLLACFLGQKVFGLDKHTSWL 136

Query: 121 VGVGSSICGSAVAATAPVIKADDEVAKAISVIFLFNMLAALLFPSLGQLLG--LSNEG 178
 +G GSSICG +AV AT PV+KA+ +V A++ + +F +A L+P++ L+ S E
 25 Sbjct: 137 IGAGSSICGAAAVLATEPVVKAESKVTVAATVVFVGTVAIFLYPAIYPLMSQWFSPET 196

Query: 179 FAIFAGTAVNDTSSVTATATAWDALHHSNTLDGATIVKLTRTLAILPITLGLSLYRAKKE 238
 F I+ G+ V++ + V A A + + A I K+ R + + P + L+ R K+
 30 Sbjct: 197 FGIYIGSTVHEVAQVVAAGHAIS----PDAENAAVISKMLRVMLAPFLILLAA-RVKQL 251

Query: 239 HDIIVTEENFSLRKSFPFRFILFFLLASLITLMTSLGVSADSFHYLKTLSKFFIVMAMAAI 298
 + E + + P F + F++ ++ + + + L TL F + MAMAA+
 35 Sbjct: 252 SGANSGEKSKI--TIPWFAILFIVVAIFNSFHL---LPQSVVNMLVTLDTFLLAMAMAAL 306

Query: 299 GLNTYNLVKLIKTTGGQAILLGAI--CWVAI 325
 GL T++ L K G + +L+ + W+ +
 40 Sbjct: 307 GLTTHVSALKKAGAKPLLALVLFALWLV 335

An alignment of the GAS and GBS proteins is shown below.

Identities = 225/333 (67%), Positives = 277/333 (82%), Gaps = 3/333 (0%)

40 Query: 11 KIPGLILCFIIAIPSWLLGLYLPLIGAPVFAILIGIIVGSFYQNRQLFNKGIAFTSKYIL 70
 K+PGL+LC ++A+P+W LG P+IGAPVFAIL+G+++ FY++R +GI+FTSKYIL
 Sbjct: 7 KLPGLLLCLLLALPAWCLGRLPFIIGAPVFAILLGMLLALFYEHRDKTKEGISFTSKYIL 66

45 Query: 71 QTAVVLLGFGLNLMQVMKVGISSLPIIIMTISISLIIAYVLQKLFKLDKTIATLIGVGSS 130
 QTAVVLLGFGLNL QVM VG+ SLPIII TI+ +L++AY LQK +LD ATL+GVGSS
 Sbjct: 67 QTAVVLLGFGLNLTQVMAVGMQSLPIIISTITATALLVAYGLQKWLRLDVNTATLIGVGSS 126

50 Query: 131 ICGGSAAIATAPVINAKDDEVAQAISVIFLFNLAALIFPTLGNFIGLSDHGFALFAGTA 190
 ICGGSAAIATAPVI AKDDEVA+AISVIFLFN+LAAL+FP+LG +GLS+ GFA+FAGTA
 Sbjct: 127 ICGGSAAIATAPVIKADDEVAKAISVIFLFNMLAALLFPSLGQLLGLSNEGFAIFAGTA 186

55 Query: 191 VNDTSSVTATATAWDALHHSNTLDGATIVKLTRTLAIIPITIVLSIYHMKQTQ---KEQS 247
 VNDTSSVTATATAWDA++HSNTL GATIVKLTRTLAI+PIT+ LS+Y K+ E++
 Sbjct: 187 VNDTSSVTATATAWDALHHSNTLDGATIVKLTRTLAILPITLGLSLYRAKKEHDIVTEEN 246

60 Query: 248 VSVTKIFPKFVLYFILASLLTTIVASLGFSRLRIFEPLKVLKSKFFIVMAMGAIGINTNVSK 307
 S+ K FP+F+L+F+LASL+TT++ SLG S F LK LSKFFIVMAM AIG+NTN+ K
 Sbjct: 247 FSLRKSFPFRFILFFLLASLITLMTSLGVSADSFHYLKTLSKFFIVMAMAAIGLNTNLVK 306

60 Query: 308 LIKTTGGKSILLGAACWLGIIVSLTMQAILGTW 340
 LIKTTGG++ILLGA CW+ I +VSL MQ LG W
 Sbjct: 307 LIKTTGGQAILLGAICWVAITLVSLLAMQLSLGIW 339

A related GBS gene <SEQ ID 8889> and protein <SEQ ID 8890> were also identified. Analysis of this protein sequence reveals the following:

-1948-

Lipop: Possible site: -1 Crend: 10
McG: Discrim Score: 22.17
GvH: Signal Score (-7.5): -0.429999
 Possible site: 41

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 8 value: -7.91 threshold: 0.0

INTEGRAL Likelihood = -7.91 Transmembrane 94 - 110 (86 - 115)

INTEGRAL Likelihood = -7.75 Transmembrane 154 - 170 (150 - 176)

INTEGRAL Likelihood = -7.11 Transmembrane 316 - 332 (312 - 339)

INTEGRAL Likelihood = -6.16 Transmembrane 258 - 274 (253 - 278)

INTEGRAL Likelihood = -2.71 Transmembrane 218 - 234 (217 - 234)

INTEGRAL Likelihood = -1.49 Transmembrane 286 - 302 (283 - 302)

INTEGRAL	Likelihood = -0.96	Transmembrane	73 - 89 (73 - 89)
----------	--------------------	---------------	--------------------

INTEGRAL Likelihood = -0.27 Transmembrane 121 - 137 (121 - 137)

PERIPHERAL Likelihood = 3.29 175

modified ALOM score: 2.08

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4163(Affirmative) < succ>

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF02059 (334 - 1284 of 1620)

EGAD|10465|EC2158(17 - 335 of 349) hypothetical 36.9 kd protein in lysP-nfo intergenic region {Escherichia coli} OMNI|NT01BC2574 conserved hypothetical protein SP|P33019|YEIH_ECOLI HYPOTHETICAL 36.9 KDA PROTEIN IN LYSP-NFO INTERGENIC REGION. GP|405879|gb|AAA60511.1|U00007 yeiH {Escherichia coli} GP|1788482|gb|AAC75219.1|AE000305 orf, hypothetical protein {Escherichia coli} PIR|E64984|E64984 hypothetical 36.9 kD protein in lysP-nfo intergenic region - Escherichia coli (strain K-12)

%Match = 12.7

%Identity = 32.3 %Similarity = 57.1

Matches = 103 Mismatches = 125 Conservative Sub.s = 79

270 300 330 360 390 435 462
 YSGPLSVFLSRFKACDIIVNVRTIMLFKEKIPGLILCFIIAIPSWLLGLYPLI-----GAPVFAILIGIIVG-SFYQN
 ||| | : | | : : | : || : | : : | : :
 MTNITLQKQHRTLWHFIPGLALSAVIT-GVALWGGSIPAVAGAGFSALTALAILGMVLGNTIYPH
 10 20 30 40 50 60

489 519 549 579 609 636 666 696
R-QLFNKGIAFTSKYILQTA VVLLGFLGNLMQVMKVGISSLP I IIMTISISLI IAYVL-QKLFKLDKTIATLIGVGSSIC
: : |:| :||: ::| || :| :| ||| :| :|:| ::::| ||| : ||| |||||
IWKSCDGGVLFQAKQYLLRLG I ILYGFRLTFSQIADVGISG I IIDLVLTSSTFLACFLGQKVFLDKHTSWLIGAGSSIC
80 90 100 110 120 130 140

726 756 786 816 840 870 900 930
GSSAIAATAPVINAKDDEVAQAISVIFLFLNILAALIFPTLGNFIG--LSDHGFALFAGTAVNDTSSVTATATAWDAINHS
| : | : || | : : | | :: : : | : | :: | : : | | : : | : | : | : | : | : | : | : | : | : | :
GAAAVLATEPVVKAEASKVTVA VATVVIFGTVAIFLYPAIYPLMSQWFSPE TFGIYGISTVHEVAQVVA--AGHAI-SP
160 170 180 190 200 210 220

960 990 1020 1050 1077 1107 1134 1164
NTLGGATIVKLTRTLAIIPITIVLSIYHMQTQKEQSVSVTKI-FPKFVLYFILASLLTTIVASLGF-SLRIFEPLKVL
:
| | : | : : | | : | : | : | : | : : : | : : :
DAENAAVISKMLRVMMLAPFLILLAA-RVKOLSGANSGEKSKITIPWFILFIVVAIF---NSFHLLPQSVMNMLVTL

230 240 250 260 270 280 290

1194 1224 1254 1284 1314 1344 1374 1404
KFFIVMAMGAIGINTNVSKLTKTGGSILLGAACWLGIIIVSLTMQAILGTW*SCLKLNICNRFHKCYNEIDIKRREHYGI
|::|||::|::||| | | :|: : :
TFLLAMAMAALGLTTHVSALKKGAKPLLMALVLFPAWLIVGGGAINYVIQSVIA

310 320 330 340

-1949-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1729

A DNA sequence (GBSx1834) was identified in *S.agalactiae* <SEQ ID 5373> which encodes the amino acid sequence <SEQ ID 5374>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL      Likelihood =-10.93      Transmembrane      7 - 23 ( 1 - 27)
```

```
----- Final Results -----
```

```
      bacterial membrane --- Certainty=0.5373(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5375> which encodes the amino acid sequence <SEQ ID 5376>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
```

```
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL      Likelihood =-16.34      Transmembrane      22 - 38 ( 13 - 42)
```

```
----- Final Results -----
```

```
      bacterial membrane --- Certainty=0.7538(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 56/215 (26%), Positives = 111/215 (51%), Gaps = 5/215 (2%)
```

```
Query: 7  VFLTIVLVLILIVGAGGLYFWNNHQSLEGKWRTVSLEKQVEKEIEQQLGSAADMGISAAD 66
      +F+ ++ LIL+   G+ + N+ S+EG WRT S+++++ + ++L           I +
Sbjct: 22  LFVFIIFLILAVLFGVRYRNS--SIEGIWRTTSIDQKLGDDFAKRLTGLHQSPILDD- 78

Query: 67  LVKGANMHMNVKNDEAKITVTAQIDEVKFHOAIKTFIDKALEKQLKDQGLTYNDLSEAGK 126
      L+ + M + VKN+ . ++ + Q++ F + + +   L K LK+ L  DLS  +
Sbjct: 79  LLTSSQMILTVKNNNVDLSPFSVQVERDIFVKRLAAYHQNELKLTLENHLVVGDLSSKER 138

Query: 127 KIFDETKITDQQIDQQIDRSFQSAQAAGGKYNTNTGEMTLPVMDGKVVHRLTSVIKV-SH 185
      +I + +      +++ +D++F+ A  GGKYN TG ++ V+ GKV+R+ I +
Sbjct: 139 QIIENSMPASHELEMILDQAFEKLASQIGGKYNQKTGHL SAVVLKGVNRILHTIDIKEE 198

Query: 186 INKKANAFYGNIVKNGEKTAYKKEGSKL-ILGNEK 219
      +      +F  ++      Y + G KL +LG+EK
Sbjct: 199 VAAGHTSFSKGLLTPNGYFDYTRFGKKLELLGDEK 233
```

SEQ ID 5374 (GBS288) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 3; MW 53.7kDa).

GBS288d was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 154 (lane 8-10; MW 26kDa) and in Figure 183 (lane 3; MW 26kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 187 (lane 11; MW 51kDa). Purified GBS288d-GST is shown in lane 8 of Figure 237.

-1950-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1730

A DNA sequence (GBSx1835) was identified in *S.agalactiae* <SEQ ID 5377> which encodes the amino acid sequence <SEQ ID 5378>. Analysis of this protein sequence reveals the following:

Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3885(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1731

A DNA sequence (GBSx1836) was identified in *S.agalactiae* <SEQ ID 5379> which encodes the amino acid sequence <SEQ ID 5380>. Analysis of this protein sequence reveals the following:

Possible site: 51
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood =-12.37 Transmembrane 67 - 83 (63 - 89)
INTEGRAL Likelihood = -3.72 Transmembrane 139 - 155 (137 - 158)
INTEGRAL Likelihood = -1.54 Transmembrane 115 - 131 (114 - 131)

----- Final Results -----

bacterial membrane --- Certainty=0.5946(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10905> which encodes amino acid sequence <SEQ ID 10906> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1732

A DNA sequence (GBSx1837) was identified in *S.agalactiae* <SEQ ID 5381> which encodes the amino acid sequence <SEQ ID 5382>. Analysis of this protein sequence reveals the following:

Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4709(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1951-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1733

A DNA sequence (GBSx1838) was identified in *S.agalactiae* <SEQ ID 5383> which encodes the amino acid sequence <SEQ ID 5384>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2191(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC98427 GB:M63481 20-kDa protein [Streptococcus sanguinis]

Identities = 119/163 (73%), Positives = 146/163 (89%)

Query: 1 MTTFLGNPVTFTGKQLQVGDIKDFLLIATDLSQKSLKDFEGKKKVISVPSIDTGICSK 60

MTTFLGNPVTFTGKQLQVGD A DF L ATDLS+K+L DF GKKKV+S++PSIDTG+CS

Sbjct: 1 MTTFLGNPVTFTGKQLQVGDTAHDFTATDLSKKTADLADFAGKKKVLSTIIPSIDTGVCS 60

Query: 61 QTRTFNEELSELNNTVTVITVMSDLPFAQKRWCSAEGLDNVILLSDFYDHSFGQYALLMN 120

QTR FN+ELS+LDNTVVITVS+DLPFAQ +WC+AEG++N ++LSD++DHSFG++YA+L+N

Sbjct: 61 QTRRFNQELSDLDNTVTVITVSDLPFAQGWCAEGIENAVMLSDYFDHSFGRDYAVLIN 120

Query: 121 EWHLLTRAVLILDEHNKVTYTEYVDNVNSDVEDYEAANAAIL 163

EWHLL RAVL+LDE+N VTY EYVDN+N++ DY+AAI A K L

Sbjct: 121 EWHLLRAVLVLDENNTVTYAEYVDNINTEPDYDAATAAVKSL 163

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1734

A DNA sequence (GBSx1839) was identified in *S.agalactiae* <SEQ ID 5385> which encodes the amino acid sequence <SEQ ID 5386>. This protein is predicted to be DNA alkylation repair enzyme. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4729(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB40581 GB:AJ010128 DNA alkylation repair enzyme [Bacillus

cereus]

Identities = 67/217 (30%), Positives = 119/217 (53%), Gaps = 5/217 (2%)

-1952-

Query: 6 SLERKFKAAADKEVSKQOEAYLRHHFKCYGIKSPERRMLYKELIKAARKQAKIDWQLLDK 65
 +L+ F A + E ++ Y+++HF GI++PERR L K++I+ + D+Q++ +
 Sbjct: 7 ALQEHFIANQNPEKAEPMARYMKNHFPFLGIQTERRQLLDKQVIQIHTLPDQKDFQVIVR 66

Query: 66 -CWQSDYREYHHFVLDYLLAMSQFLTYNDCSRLEFYARHQWWSIDVLTQIF-GNLSLK 123
 W RE+ LD + + LE + WWD++D + F GN+ L+
 Sbjct: 67 ELWDLPEREFQAAALDMMQKYKMHINETHIPFLEELIVTKSWWDTVDSIVPTFLGNIFLQ 126

Query: 124 DDKVMNL-LSEWSLDQDFWMRLAIEHQLGFKETNTDILSLFILRNTGSQEFFINKAIG 182
 ++++ + +W + W++R AI QL +K+K + ++L I + S+EFFI KAIG
 Sbjct: 127 HPELISAYIPKWIASDNIWLQRAILFQLKYKQKMDDELLFWVIGQLHSSKEFFIQKAIG 186

Query: 183 WALRDYSKYNKVWVKDFISNHCDELSTLSIREGSKYL 219
 W LR+Y+K V +++ N +EL+ LS RE K++
 Sbjct: 187 WVLREYAKTKSDVVWEYVQN--NELAPLSRREAIKHI 221

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1735

A DNA sequence (GBSx1841) was identified in *S.agalactiae* <SEQ ID 5387> which encodes the amino acid sequence <SEQ ID 5388>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2117(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA81648 GB:Z27121 unknown [Mycoplasma hominis]
 Identities = 67/281 (23%), Positives = 113/281 (39%), Gaps = 52/281 (18%)

Query: 3 FVFDFIDGTLCDGMS--LSKEIQGILERAQIDYGHVRTFATARSYRDTIGILGDKLSLSK 60
 F D+DGTL D + + + + +++A + GH V+ T R +R T+ + +KL L+
 Sbjct: 14 FAIDLDTLLADSANGTVHPKTEEAIIKA-VAQGHIVSIITGRPWRSTLPVY-EKLGLNA 71

Query: 61 IIG-LNGATLHENGHLVDSYYLQSDFFSTIISYCHRHQIPYFVD-----EVFNATYQA 113
 I+G NGA +H FF I+Y +++ Y + E+ NYA
 Sbjct: 72 IVGNYNGAHIHNPA-----DPFFIPAITYLDLNEVLYILGDEKVKKEITNYAIEGP 122

Query: 114 SKIPFIAYVDPQ-----KRGELLEVSKE-----KPIKMVLYFGDQLGR 152
 + + + DP K E + + KI KP VL L R
 Sbjct: 123 DWVQLM-HRDPNLERVFGFNQATKFRECINLEKIPKPTGIVFDVKPDTDVLELLTYLKR 181

Query: 153 ADQMLAELNRFGLSSHHFFHEFEKCLYINPIAVDKGKATKKLFG-----NRFIAFGNDKN 206
 L E + + F+ I I +DKGK + + +A G+ N
 Sbjct: 182 RYGDLDGEFSSWSKGEGLSPVFD---ITSIGIDKGKVISLIMRYYNIDDDTVAMGDSYN 237

Query: 207 DISMFDAAHYSVQVGDFDELTPYANLRVSRESVHEGITTLE 247
 D+SM++ A+ V + + L + V +++ EG F
 Sbjct: 238 DLSMYNVANVCVSPANAEPLIKMSTVVMKQTNKEGAVGYF 278

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1953-

Example 1736

A DNA sequence (GBSx1842) was identified in *S.agalactiae* <SEQ ID 5389> which encodes the amino acid sequence <SEQ ID 5390>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2383(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB90005 GB:AE001018 A. fulgidus predicted coding region AF1244

[Archaeoglobus fulgidus]

Identities = 22/48 (45%), Positives = 35/48 (72%)

Query: 150 GKSIGELNVWHQTGATIVAIEHEGKFIVSPGPFVSVIEQGDHIFVVGDE 197

GKSIGEL + +TGAT++A+ + K I+SP P +V+E GD + +G++

Sbjct: 102 GKSIGELGIRSKTGATVIAVLKKEKTIISPSPETVLEPGDKVVVIGEK 149

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5391> which encodes the amino acid sequence <SEQ ID 5392>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2446(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 163/213 (76%), Positives = 196/213 (91%)

Query: 1 MVSEQSEIVTSKYQKIAVAVAQRIANGDYEVGEKLSRTTTIASTFNVSPETARKGLNILA 60

++S + EI +SKYQKIA++VAQRIANG+YEVGEKLSRTTTIASTFNVSPETARKGLNILA

Sbjct: 1 VISPKKEITSSKYQKIAISVAQRIANGYEVGEKLSRTTTIASTFNVSPETARKGLNILA 60

Query: 61 DLQILTLKHGSGAIIISKEKAIEFLNQYETSHSVAILKGKIRDNIKAQQQEMEELATLVD 120

DL+ILTLKHGSGAI+LSKE+AIEF+NQYE++HS+A+LK KIR+ I Q + ME++A LV+

Sbjct: 61 DLKILTLKHGSGAIVLSKERAIEFINQYESTHSIAVLKEKIRETINDQGKAMEKMAVLVN 120

Query: 121 DFLQLQTRAVSKQYPLAPYEIIIVSEDEHLGKSIGELNVWHQTGATIVAIEHEGKFIVSPG 180

DFL+Q+++VSKQYPLAPYEII ++DSEH GKSIG LN+WHQTGATIVAIEH G+FIVSPG

Sbjct: 121 DFLMQSQSVSKQYPLAPYEIICNDSEHFGKSIGVLNIWHQTGATIVAIEHAGQFIVSPG 180

Query: 181 PFSVIEQGDHIFVVGDEDEVYARMKTYFNLRMGL 213

P+SVIE+GDHI+FVGDE V +RMKT+FNLR GL

Sbjct: 181 PYSVIEKGDHIYFVGDESVISRMTFFNLRKGL 213

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1737

A DNA sequence (GBSx1844) was identified in *S.agalactiae* <SEQ ID 5393> which encodes the amino acid sequence <SEQ ID 5394>. This protein is predicted to be gls24. Analysis of this protein sequence reveals the following:

Possible site: 16

-1954-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2855(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9361> which encodes amino acid sequence <SEQ ID 9362> was also identified.

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA86383 GB:U23376 putative 20-kDa protein [Lactococcus lactis]
 Identities = 63/124 (50%), Positives = 84/124 (66%)

15 Query: 1 MSGGFFSNLKNSVNSDSVTGDNVEVGKTEVAVDLDIVVEYGKDIPAIVESIKAIVSQN 60
 + GGFFSNL ++N+D VT GV+VEVG +VAVDL +V EY K++P I E IK ++ +
 Sbjct: 55 VEGGFFSNLTGKLINTDDVITGVDVEVGKTOVAVDLKVVEYRKNVDPDIYEKIKEVIRKE 114

20 Query: 61 VEVMTHLKVVELNANVVDIKTKAEHEADSVTVQDRVSDAAQATGNFASEQAGKAKAAISS 120
 V MT L+VVE+N V DIKT + + D V++QDRV+ AAQ TG F SEQ K K +
 Sbjct: 115 VAAMTELEVVEVNVITVDIKTKEQKEDDVSIQDRVTSAAQTGKFTSEQVDKVKDKVED 174

25 Query: 121 GAEK 124
 +K
 Sbjct: 175 NTDK 178

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5395> which encodes the amino acid sequence <SEQ ID 5396>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2534(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 94/137 (68%), Positives = 108/137 (78%), Gaps = 8/137 (5%)

40 Query: 1 MSGGFFSNLKNSVNSDSVTGDNVEVGKTEVAVDLDIVVEYGKDIPAIVESIKAIVSQN 60
 ++GGFFSN+KN++VNS+SVTDGV+VEVG+KEVAVDL I+VEYGKDIPAI ESIKAIVSQN
 Sbjct: 35 VTGGFFSNIKNNLVNSESVTGDSVEVGSKEVAVDLAIIVEYGKDIPAIAESIKAIVSQN 94

45 Query: 61 VEVMTHLKVVELNANVVDIKTKAEHEADSVTVQDRVSDAAQATGNFASEQAGKAKAAISS 120
 V+ MTHLKVVE+N NVVDI+TK EHEA SVTVQDRV+ AA +T F SEQ K K IS
 Sbjct: 95 VDSMTHLKVVEVNVNVVDIRTKEEHEAASVTQDRVTSAASTSQFVSEQTEKLKDTISD 154

50 Query: 121 GAEKTKEAVSNGTEAAK 137
 N EAAK
 Sbjct: 155 -----TVNSDEAAK 163

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1738

55 A DNA sequence (GBSx1845) was identified in *S.agalactiae* <SEQ ID 5397> which encodes the amino acid sequence <SEQ ID 5398>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

-1955-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3393(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

Example 1739

A DNA sequence (GBSx1846) was identified in *S.agalactiae* <SEQ ID 5399> which encodes the amino acid sequence <SEQ ID 5400>. Analysis of this protein sequence reveals the following:

15 Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.3168(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

Example 1740

A DNA sequence (GBSx1847) was identified in *S.agalactiae* <SEQ ID 5401> which encodes the amino acid sequence <SEQ ID 5402>. This protein is predicted to be gls24. Analysis of this protein sequence reveals the following:

30 Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2718(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:AAA86383 GB:U23376 putative 20-kDa protein [Lactococcus lactis]
 Identities = 95/157 (60%), Positives = 120/157 (75%)

 Query: 18 VRGELTFEDKIVIEKIVGIAIEHVDGLLAVNGGFFSNLKNVSVNSDSVTDGVNVEVGKKQV 77
 ++G LT+EDKV++KIVG+A+E VDGLL+V GGFFSNL ++N+D VT GV+VEVGK QV
 Sbjct: 27 IKGALTYEDKVVQKIVGLALESVVDGLLSVEGGFFSNLTGKLIINTDDVTGVDVEVGKTQV 86

 Query: 78 AVDLDIVAEYQKHVPTIFADIKKVVEAEVKRMTDLEVVEVNVNVVDIKTRAQHEEDSVTL 137
 AVDL +V EY+K+VP I+ IK+V+ EV MT+LEVVEVNV V DIKT+ Q +ED V++
 Sbjct: 87 AVDLKVVTEYRKNVDPDIYEKIKEVIRKEVAAMTELEVVEVNVTVTDIKTKEQQKEDDVSI 146

 Query: 138 QDRVTSAAQATGEFASNQVSNVKSAVGSGVDKVEDMK 174
 QDRVTSAAQ TG+F S QV VK V DK +K

-1956-

Sbjct: 147 QDRVTSAAQTGKFTSEQVDKVKDKVEDNTDKEARVK 183

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5403> which encodes the amino acid sequence <SEQ ID 5404>. Analysis of this protein sequence reveals the following:

5 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3896(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 123/180 (68%), Positives = 158/180 (87%), Gaps = 1/180 (0%)

Query: 1 MTETYIKNTTNNSGTTAVRGELTFEDKVIEKIVGIAIEHVDGLLAVNGGFFSNLKNSVVN 60
 MTETYIKNT+ + T+A+RG+LT++DKVIEKIVG+A+E+VDGLL VNGGFF+NLK+ +VN
 Sbjct: 1 MTETYIKNTSKDL-TSAIRGQLTYDDKVIEKIVGLALENVVDGLLGVNGGFFANLKDCLVN 59

20 Query: 61 SDSVTDGVNVEVGKKQVAVDLDIVA EYQKHVPTIFADIKKVEAEVKRMTDLEVVEVNVN 120
 ++SV DGVNVEVGKKQVAVDLDIVA EYQKHVPTI+ IK +VE EVKRM TDL+V+EVNV
 Sbjct: 60 TESVRDGVNVEVGKKQVAVDLDIVA EYQKHVPTIYDSIKSIVEEEVKRMTDLDVIEVNVK 119

25 Query: 121 VVDIKTRAQHEEDSVTLQDRVTSAAQATGEFASNQVSNVKS AVSGVDKVEDMKSEPRVQ 180
 VVDIKT+ Q E + V+LQD+V+ A++T EF S+QV NVK++V +GV+K++D K+EPRV+
 Sbjct: 120 VVDIKTKEQFEAEKVSLQDKVSDMARSTSEFTSHQVENVKASVDNGVEKLQDQKAEPRVK 179

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 1741

A DNA sequence (GBSx1848) was identified in *S.galactiae* <SEQ ID 5405> which encodes the amino acid sequence <SEQ ID 5406>. This protein is predicted to be a 6-kDa protein. Analysis of this protein sequence reveals the following:

35 Possible site: 22
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -9.29 Transmembrane 25 - 41 (23 - 52)

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:AAA86382 GB:U23376 putative 6-kDa protein [Lactococcus lactis]
 Identities = 27/61 (44%), Positives = 45/61 (73%)

 Query: 3 EFVRKYRYPLGGAVIGLVLAAMIVTIGFFKTILALVIIVLGAYAGLYVQRTGMLDQFFNK 62
 ++ K RYP+ G ++G ++A I TIGF+K IL L +I LG Y GL+++++G++DQF N+
 Sbjct: 2 DYFEKNRYPIIGGIVGALIAVCIFTIGFWKMILVFLIGLGIYIGLFLKKSGIIDQFINR 61

50 Query: 63 R 63
 +
 Sbjct: 62 K 62

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5407> which encodes the amino acid sequence <SEQ ID 5408>. Analysis of this protein sequence reveals the following:

-1957-

Possible site: 28

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -11.73 Transmembrane 11 - 27 (6 - 50)
 INTEGRAL Likelihood = -7.11 Transmembrane 33 - 49 (27 - 50)

----- Final Results -----

bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 28/61 (45%), Positives = 48/61 (77%)

Query: 3 EFVRKYRYPLGGAVIGLVLAAMIVTIGFFKTILALVIIIVLGAYAGLYVQRTGMLDQFFNKR 63

EF K++YP+ G ++GL++A +++ G FKT+LA++ I+LG Y GLY ++TG++DQF N++

Sbjct: 2 EFYEKFKYPIIGGLVGLIIAILLMAFGLFKTLAIIFIILGIYGGLYAKKTGIIDQFLNRK 62

A related GBS gene <SEQ ID 8891> and protein <SEQ ID 8892> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8

McG: Discrim Score: 12.56

GvH: Signal Score (-7.5): -1.11

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 1 value: -9.29 threshold: 0.0

INTEGRAL Likelihood = -9.29 Transmembrane 25 - 41 (23 - 52)

PERIPHERAL Likelihood = 12.25 44

modified ALOM score: 2.36

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

44.3/73.8% over 60aa

Lactococcus lactis

EGAD|42618| putative 6-kDa protein Insert characterized

GP|727435|gb|AAA86382.1||U23376 putative 6-kDa protein Insert characterized

ORF01006(307 - 489 of 792)

EGAD|42618|45008(2 - 62 of 62) putative 6-kDa protein {Lactococcus

lactis}GP|727435|gb|AAA86382.1||U23376 putative 6-kDa protein {La

ctococcus lactis}

%Match = 11.6

%Identity = 44.3 %Similarity = 73.8

Matches = 27 Mismatches = 16 Conservative Sub.s = 18

159 189 219 249 279 309 339 369
 TNVPEQLEHIQSDVELGLKEFFGLEKKMNTRVFVKQVEEENVGNAKTNKSRVE*ESNMSEFVRKYRYPLGGAVIGLVLA

:: | |||: | ::| ::|
 MDYFEKNRYPIIGGIVGALIAV

10 20

399 429 459 489 519 549 579 609
 MIVTIGFFKTILALVIIIVLGAYAGLYVQRTGMLDQFFNKRK*NFSFIFILHYLNKRKRNYD*NLHQKH*NQFWHDSCSW

| |||:| || | :| || | ||:::|:| ||| |::
 CIFTIGFWKMIIVLFLIGLGIYIGLFLKKSGLIDQFINRK

40 50 60

-1958-

SEQ ID 5406 (GBS14) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 9 (lane 4; MW 33.3kDa). The GBS14-GST fusion product was purified (Figure 190, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 263), which confirmed that the protein is immunoaccessible on GBS bacteria.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1742

A DNA sequence (GBSx1849) was identified in *S.galactiae* <SEQ ID 5409> which encodes the amino acid sequence <SEQ ID 5410>. Analysis of this protein sequence reveals the following:

```

10  Possible site: 27
    >>> Seems to have an uncleavable N-term signal seq
        INTEGRAL    Likelihood = -18.63    Transmembrane    61 - 77 ( 51 - 83)
        INTEGRAL    Likelihood = -7.91     Transmembrane    10 - 26 ( 7 - 28)

15  ----- Final Results -----
        bacterial membrane --- Certainty=0.8451(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

- 20 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5411> which encodes the amino acid sequence <SEQ ID 5412>. Analysis of this protein sequence reveals the following:

```

    Possible site: 29

25  >>> Seems to have a cleavable N-term signal seq.
        INTEGRAL    Likelihood = -16.19    Transmembrane    71 - 87 ( 63 - 93)

    ----- Final Results -----
        bacterial membrane --- Certainty=0.7474(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
30  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```

35  Identities = 87/193 (45%), Positives = 127/193 (65%), Gaps = 4/193 (2%)

Query: 1  MSKGLKSLYTLGLISLTLGLFVAVISKQHIYLP-SFNWLDWDFN-LPSPIDVGMHYHYFF 58
      MSK LK Y L+GL+ L++ G+V I+ +IYLP S+ WL W + P+ +D + +Y+F
Sbjct: 9  MSKLLKISYCLVGLVLLSVFGWVVGITGGYIYLPYSYRWLSWGMDSPFNLLDSALSYYYF 68

40  Query: 59  WGALVLFVIVLLAILVVLFPYPRRYTEYKLA--DKTGKMLKKS AIEGFVKTEVLKTGLMK 116
      W ALVLFVI LA+LV++ YPR YTE +L +K G L+LKKS AIE +V T + GLM
Sbjct: 69  WTALVLFVITFLALLVILYIPRIYTEVQLRHKNKKGTL L LKKS AIESYVATAIQTAGLMP 128

45  Query: 117  SPSVTAHLYKKKVKVDVKGLLTSRTNVPEQLEHIQSDVELGLKEFFGLEKKMNTRVFVKQ 176
      +P+VTA LYK+K + VKG L SR V +Q+ ++ +E GL EFG+ +N +V+VK
Sbjct: 129  NPTVTAKLYKRKFNIIVKGRLASRVAVADQISGVKEGIEKGLTEFFGINYPVNFVKVYVKD 188

Query: 177  VEEENVGNAKTNK 189
      + + + + N+
50  Sbjct: 189  IADSDRKHITRNR 201

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1959-

Example 1743

A DNA sequence (GBSx1850) was identified in *S.agalactiae* <SEQ ID 5413> which encodes the amino acid sequence <SEQ ID 5414>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
5  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -9.82    Transmembrane    56 - 72 ( 52 - 81)
    INTEGRAL    Likelihood = -6.42    Transmembrane     4 - 20 (  1 - 23)

----- Final Results -----
10      bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:CAB12244 GB:Z99106 similar to hypothetical proteins from B. subtilis [Bacillus
    subtilis]
    Identities = 31/76 (40%), Positives = 48/76 (62%)

Query: 1  MSLIWSLIVGAIIGAIAGAVTNKGGSMGWIANILAGLVGSFVGQSLLGTWGP KLAGMALI 60
20      +S + SL+V +IG I A+      G +++AGL+G+++G LLGTWGP LAG A+
Sbjct: 2  LSFLVSLVVAIVIGLIGSAIVGNRLPGGIFGSMIAGLIGAWIGHGLLGTWGPSLAGFAIF 61

Query: 61 PSIVGAIIVVIVTSFV 76
      P+I+GA I V + +
25  Sbjct: 62 PAIIGAAIFVFLGLI 77

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5415> which encodes the amino acid sequence <SEQ ID 5416>. Analysis of this protein sequence reveals the following:

```

Possible site: 55
30  >>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -7.59    Transmembrane    60 - 76 ( 56 - 80)

----- Final Results -----
35      bacterial membrane --- Certainty=0.4036(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

40  >GP:CAB12244 GB:Z99106 similar to hypothetical proteins from B. subtilis [Bacillus
    subtilis]
    Identities = 28/76 (36%), Positives = 47/76 (61%)

Query: 1  MGLIWTLIVGALIGVIAGALT KKGSMGWIANIAAGLVGS SVGQALLGSWG PSLAGMSLI 60
45      + + +L+V +IG+I A+      G ++ AGL+G+ +G LLG+WGPSLAG ++
Sbjct: 2  LSFLVSLVVAIVIGLIGSAIVGNRLPGGIFGSMIAGLIGAWIGHGLLGTWGPSLAGFAIF 61

Query: 61 PSVIGAVIVMITSFV 76
      P++IGA I V + +
50  Sbjct: 62 PAIIGAAIFVFLGLI 77

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 63/82 (76%), Positives = 74/82 (89%)

55  Query: 1  MSLIWSLIVGAIIGAIAGAVTNKGGSMGWIANILAGLVGSFVGQSLLGTWGP KLAGMALI 60
      M LIW+LIVGA+IG IAGA+T KGGSMGWIANI AGLVGS VQQ+LLG+WGP LAGM+LI
Sbjct: 1  MGLIWTLIVGALIGVIAGALT KKGSMGWIANIAAGLVGS SVGQALLGSWG PSLAGMSLI 60

Query: 61 PSIVGAIIVVIVTSFVLGKMNN 82
      PS++GA+IVV++TSFVL K NN
60  Sbjct: 61 PSVIGAVIVMITSFVLNKTNN 82

```

-1960-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1744

- 5 A DNA sequence (GBSx1851) was identified in *S.agalactiae* <SEQ ID 5417> which encodes the amino acid sequence <SEQ ID 5418>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

10 INTEGRAL Likelihood = -9.82 Transmembrane 88 - 104 (84 - 111)
INTEGRAL Likelihood = -8.07 Transmembrane 29 - 45 (27 - 54)

----- Final Results -----

15 bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:CAB12244 GB:Z99106 similar to hypothetical proteins from B. subtilis [Bacillus subtilis]
Identities = 29/77 (37%), Positives = 47/77 (60%)

Query: 31 IMGLIWSLIVGAIIGAIAGAITNKGSMGWIANILAGLVGSFVGQSL LGTWGPKLADMAL 90
++ + SL+V +IG I AI G +++AGL+G+++G LLGTWGP LA A+
25 Sbjct: 1 MLSFLVSLVVAIVIGLIGSAIVGNRLPGGIFGSMIAGLIGAWIGHGLLGTWGPSLAGFAI 60

Query: 91 IPSIVGAIIVIIIVTSFV 107

P+I+GA I + + +

Sbjct: 61 FPAIIGAAIFVFLGLI 77

- 30 There is also homology to SEQ ID 5416:

Identities = 60/79 (75%), Positives = 72/79 (90%)

Query: 32 MGLIWSLIVGAIIGAIAGAITNKGSMGWIANILAGLVGSFVGQSL LGTWGPKLADMALI 91
MGLIW+LIVGA+IG IAGA+T KGGSMGWIANI AGLVGS VGQ+LLG+WGP LA M+LI
35 Sbjct: 1 MGLIWTLIVGALIGVIAGALT KKGSMGWIANIAAGLVGS SVGQALLG SWGPSLAGMSLI 60

Query: 92 PSIVGAIIVIIIVTSFVLGK 110

PS++GA+IV+++TSFVL K

Sbjct: 61 PSVIGAVIVMITSFVLNK 79

40

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1745

- 45 A DNA sequence (GBSx1852) was identified in *S.agalactiae* <SEQ ID 5419> which encodes the amino acid sequence <SEQ ID 5420>. This protein is predicted to be ATP-dependent DNA helicase Rep (uvrD). Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1364(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1961-

A related GBS nucleic acid sequence <SEQ ID 9863> which encodes amino acid sequence <SEQ ID 9864> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAD51119 GB:AF176554 DNA helicase PcrA [Leuconostoc citreum]
Identities = 414/764 (54%), Positives = 537/764 (70%), Gaps = 23/764 (3%)

Query: 6 VEMNPLIIGMNDKQAEAVQTTDGPLLIMAGAGSGKTRVLTHRIAYLIDEKYVNPWNILAI 65
+ + L GMN+KQAEAVQTT+GPLLIMAGAGSGKTRVLTHRIA+L+ + V PW ILAI
Sbjct: 1 MSVETLTINGMNNKQAEAVQTTEGPLLIMAGAGSGKTRVLTHRIAHVQDLNVFPWRILAI 60

10 Query: 66 TFTNKAAREMRERIAL--NPATQDTLIATFHSMCVRLRREADYIGYNNFTIVDPGEQ 123
TFTNKAAREMRER AL +D ++TFH++ VRILRR+ + IG +NFTI+D Q
Sbjct: 61 TFTNKAAREMRERIAALLSEDVARDI WVSTFHALAVRILRRDGEAIGLAKNFTIIDTSAQ 120

15 Query: 124 RTLKRIIKQLNLDTKKWNERSILGTISNAKNDLLDEIAYEKQAGDMYTQVIKCYKAYQ 183
RTLKMR+I LNLDT +++ R+ILG ISNAKND+L Y K A + + + +A+ Y AYQ
Sbjct: 121 RTLKRVINDLNLDTNQYDPTILGMISNAKNDMLQPRDYAKAADNAFQETVAEVYTAYQ 180

20 Query: 184 EELRRSEAMDFDDLIMTLRLFDQNKDVLAYYQQRYQYIHVDEYQDTNHAQYQLVKLLAS 243
EL+RS+++DFDDLIM+T+ LF DVL A YQQ+++Y+HVDEYQDTN AQY +V LLA
Sbjct: 181 AELKRSQSVDFDDLIMLTIDLFQSAPDVLARYQQQFEYLHVDEYQDTNDAQYTIVNLLAQ 240

25 Query: 244 RFKNICVVGADQSIYGWRGADMQNILD FEKDYPQAKVVLLEENYRSTKKILQAANNVIN 303
R KN+ VVGADQSIYGWRGA+M NIL+FEKDYP A V+LE+NYRST+ IL AAN VIN
Sbjct: 241 RSKNLAVVGADQSIYGWRGANMNNILNFEKDYPNAHTVMLEQNYRSTQNILDAANAVIN 300

30 Query: 304 HNKNNRPKKLWTONDEGEQIVYHRANNEQEEAVFVASTIDNIVREQGNFKDFAVLYRTN 363
HN R PPKLWT+N +G+QI Y+RA E +EA F+ S I + + + DFAVLYRTN
Sbjct: 301 HNNERVPPKKLWTENGKGDQITYYRAQTEHDEANFILSNIQQLRETKHMAYSDFAVLYRTN 360

35 Query: 364 AQSRTIEEALLKSNIPYTMVGGTKFYSRKEIRDVIAYLNILANTSDNISFERIVNEPKRG 423
AQSRIEE+L+K+N+PY+MVGG KFY RKEI D++AY++++ N DN +FER+VNEPKRG
Sbjct: 361 AQSRIEESLVKANMPYSVMVGGHKFYERKEILDIMAYMSLITNPDNAAFERVVNEPKRG 420

40 Query: 424 VGPGLTEKIRSFAYEQSMSLLDASSNVMMSP-LKGKAAQAVWDLANLILTLRSNLDLSLTV 482
+G +L ++R A ++S + A ++ ++P + KAA A ++ LR + L V
Sbjct: 421 LGATSLTRLRELANRLNVSYMKAIGSIELAPSITTKAASKFLTFAEMMHLRQQSEFLNV 480

45 Query: 483 TEITENLLDKTGYLEALQVQNTLESQARIENIEEFLSVTKNFDDNPEITVEGETGLDRLS 542
TE+TE ++ ++GY + L +N +SQAR+EN+EEFLSVTK FDD + E +D ++
Sbjct: 481 TELTELVMTQSGYRQMLAEKNDPDSQARLENLEEFLSVTKFDD--KYQPEDPESIDPVT 538

50 Query: 543 RFLNDLALIADTDDSATETAEVTLMTLHAAKGLEFPVVFLLGMEGVFPLSRAIEDADEL 602
FL AL++D DD VTLMTLHAAKGLEFPVVFLLG++EG+FPLSRA+ D D L
Sbjct: 539 DFLGTTALMSDLDDFEEGDGAVTLMTLHAAKGLEFPVVFLLGLKEGIFPLSRAMMEDLL 598

55 Query: 603 EEERRLAYVGITRAEQILFLTNANTRTLFGKTSYNRPTRFIREIDDELIQ--YQGLARPV 660
EEERRLAYVGITRA + LFLTNA +R L+G+T N P+RFI EI EL++ Y GL+R
Sbjct: 599 EEERRLAYVGITRAMKKLFLTNFSLRLYGRQANEPSRFIAEISPELLETAYSGLSRDK 658

Query: 661 NSSFGVKYSKEOPTQFGQMSLQALQARKSNSQSQVTAQLQALN-ANNSHETSWEIGDV 719
+ + ++ R + + Q T + N +TSW GD
Sbjct: 659 TQKKTLFPDRK-----MQRATATTYQATPVTKITNGVTGGDQTSWSTGDK 703

Query: 720 ATHKKWGDGTVLEVSGSGKTQELKINFPGLGLKLLASVAPISK 763
+HKKWG GTV+ VSG QELK+ FP G+K+LLA+ API K
Sbjct: 704 VSHKKWGVGTVISVSGRADDQELKVAFPSEGVKQLLAFAPIQK 747

60 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5421> which encodes the amino acid sequence <SEQ ID 5422>. Analysis of this protein sequence reveals the following:

Possible site: 15
>>> Seems to have no N-terminal signal sequence

-1962-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0214(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 622/772 (80%), Positives = 699/772 (89%), Gaps = 15/772 (1%)

10

Query: 8 MNPLIIGMNDKQAEAVQTTDGPLLIMAGAGSGKTRVLTHRIAYLIDEKYVNPWNILAITF 67
 MNPL+ GMND+QA+AVQTT+GPLLIMAGAGSGKTRVLTHRIAYLIDEK+VNPWNILAITF
 Sbjct: 1 MNPLNGMNDRQAQAVQTTTEGPLLIMAGAGSGKTRVLTHRIAYLIDEKFVNPWNILAITF 60

15

Query: 68 TNKAAREMRERAIALNPATQDTLIATFHSMCVRIILRREADYIGYNRNFTIVDPGEQRTLM 127
 TNKAAREM+ERA+ALNPAT+DTLIATFHSMCVRIILRREAD+IGYNRNFTIVDPGEQRTLM
 Sbjct: 61 TNKAAREMKERALALNPATKDTLIATFHSMCVRIILRREADHIGYNRNFTIVDPGEQRTLM 120

20

Query: 128 KRIIKQLNLDTKKWNERSILGTISNAKNDLLDEIAYEKQAGDMYTQVIKCYKAYQEELR 187
 KRI+KQLN+D KKNERSILGTISNAKNDLLDE YE QA DMY+Q++A+CYKAYQEELR
 Sbjct: 121 KRILKQLNIDPKWNERSILGTISNAKNDLLDEKGYEAQAADMYSQIVARCYKAYQEELR 180

25

Query: 188 RSEAMDFDDLIMMTLRLFDQNKDVLAYYQORYQYIHVDEYQDTNHAQYQLVLLASRFKN 247
 RSEA+DFDDLIMMTLRLFD N DVLAYYQORYQYIHVDEYQDTNHAQYQL+KLLASRFKN
 Sbjct: 181 RSEALDFDDLIMMTLRLFDANPDVLAYYQORYQYIHVDEYQDTNHAQYQLIKLLASRFKN 240

30

Query: 248 ICVVGADQSIYGWRGADMQNILDFEKDYPQAKVLLLEENYRSTKKILQAANNVINHNKN 307
 ICVVGADQSIYGWRGADMQNILDFEKDYP AKVLLLEENYRSTKKILQAAN+VIN+N+N
 Sbjct: 241 ICVVGADQSIYGWRGADMQNILDFEKDYPDAKVLLLEENYRSTKKILQAANDVINNNRN 300

35

Query: 308 RRPKKLWTONDEGEQIVYHRANNEQEEAVFVASTIDNIVREQGNFKDFAVLYRTNAQSR 367
 RRPKKLWTON +GEQ+VY+RAN+E++EAVFVASTI N+ +E GKNFKDFAVLYRTNAQSR
 Sbjct: 301 RRPKKLWTONADGEQLVYYRANDEDEAVFVASTISNMSQELGKNFKDFAVLYRTNAQSR 360

40

Query: 368 TIEEALLKSNIPYTMVGGTKFYSRKEIRDVIAYLNILANTSDNISFERIVNEPKRGVGP 427
 TIEEALLKSNIPYTMVGGTKFYSRKEIRD+IAYL I+AN +DNISFERIVNEPKRGVGP
 Sbjct: 361 TIEEALLKSNIPYTMVGGTKFYSRKEIRDLIAYLTIVANPADNISFERIVNEPKRGVGP 420

45

Query: 428 TLEKIRSFAYEQSMSLLDASSNVMMSPLKGAQAQVWDLANLILTLRSNLDLSLTVTEITE 487
 TL+K+R FAYE SLL+A+SN++MSPLKGAQA+ DLAN++ LR +LD +++T++ E
 Sbjct: 421 TLDKLRQFAYESDQSLLEAASNLLMSPLKGAQAQIMDLANILQQLRQDLQMSITDLAE 480

50

Query: 488 NLLDKTGYLEALQVQNTLESQARIENIEEFLSVTKNFDNPEITVEGETGLDRLSRFLND 547
 LL+KTGYL++L++QNTLESQARIENIEEFLSVTKNFD++ E ETG+DRL RFLND
 Sbjct: 481 ALLEKTGYLDSLRLQNTLESQARIENIEEFLSVTKNFDSSASQEEDETGVDRIGRFLND 540

55

Query: 548 LALIADTDDSATETAETVTLMTLHAAKGLEFPVVFLIGMEEGVFPLSRAIEDADELEEERR 607
 LALIADTDDS E AEVTLMTLHAAKGLEFPVVFLIGMEEGVFPLSRA ED DELEEERR
 Sbjct: 541 LALIADTDDSQAEAAEVTLMTLHAAKGLEFPVVFLIGMEEGVFPLSRASEDPDELEEERR 600

Query: 608 LAYVGITRAEQILFLTNANTRTLFGKTSYNRPTRFIREIDDELIQYQGLARPVNSSFGVK 667
 LAYVGITRAE++LF+TNANTRTLFGK+SYNRPTRF++EI +EL+ Y+GLARP SSFGV+
 Sbjct: 601 LAYVGITRAEEVLFMTNANTRTLFGKSSYNRPTRFLKEISEELLSYKGLARPAQSSFGVR 660

Query: 668 YSKEQPTQFGQMSLQOALQARKSNSQSQVTAQ-LQA-----LNANNS-HET 712
 +S E TQFGQMSL +ALQARK+ +Q + +AQ +QA +N+S E
 Sbjct: 661 FSTETHTQFGQMSLSEALQARKAQAQVRQSAQPMQAHTIPSASTSSVLFPFGSNSSVEEV 720

60

Query: 713 SWEIGDVATHKKWGDGTVLEVSGSGKTQELKINFPGIGLKKLLASVAPISK 764
 +W+IGD+A HKKWGDGTVLEVSGSGKT ELKI FP +GLKKLLASVAPI KK
 Sbjct: 721 TWQIGDIAHKKWGDGTVLEVSGSGKTMELKIKFPEVGLKKLLASVAPIEKK 772

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1963-

Example 1746

A DNA sequence (GBSx1853) was identified in *S.agalactiae* <SEQ ID 5423> which encodes the amino acid sequence <SEQ ID 5424>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4741(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA88579 GB:M14339 unknown [Streptococcus pneumoniae]

Identities = 43/57 (75%), Positives = 50/57 (87%)

Query: 41 AHGGYLFITLCDQVSGLVIAISTGYEAVTLQSNINYLKAGRLDDLLTVIGTCVHNGRTT 97

AHGGYLFITLCDQ+SGLV IS G + VTLQS+INYL+AG+LDD+LT+ G CVH GRTT

Sbjct: 1 AHGGYLFITLCDQISGLVVISLGLDGVTLQSSINYLKAGKLDVLTIKGECVHQGRRTT 57

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5425> which encodes the amino acid sequence <SEQ ID 5426>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1210(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 57/97 (58%), Positives = 74/97 (75%)

Query: 2 KFNLEQVKVFENYEIENWEEGQVTLTKVVDSSLNYYGNARGGYLFITLCDQVSGLVIAIST 61

+ L + +F+NY+IE E+G + L+T+V +++LNYYGNAHGGYLFITLCDQV GLVA +T

Sbjct: 7 EMTLNVISIFDNYQIELAEKGHLILSTEVTTETALNYYGNARGGYLFITLCDQVGGGLVARTT 66

Query: 62 GYEAVTLQSNINYLKAGRLDDLLTVIGTCVHNGRTTK 98

G E+VTLQ+N NYL+AG D L V G VH GRTT+

Sbjct: 67 GVESVTLQANANYLKAGHKGDKLMVEGRLVHGGRTTQ 103

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1747

A DNA sequence (GBSx1854) was identified in *S.agalactiae* <SEQ ID 5427> which encodes the amino acid sequence <SEQ ID 5428>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3187(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-1964-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1748

A DNA sequence (GBSx1855) was identified in *S.agalactiae* <SEQ ID 5429> which encodes the amino acid sequence <SEQ ID 5430>. This protein is predicted to be uracil permease (uraA). Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have no N-terminal signal sequence

10	INTEGRAL	Likelihood = -8.65	Transmembrane	122 - 138 (117 - 146)
	INTEGRAL	Likelihood = -8.65	Transmembrane	212 - 228 (204 - 233)
	INTEGRAL	Likelihood = -7.32	Transmembrane	60 - 76 (49 - 80)
	INTEGRAL	Likelihood = -6.53	Transmembrane	149 - 165 (145 - 172)
	INTEGRAL	Likelihood = -6.48	Transmembrane	402 - 418 (401 - 420)
15	INTEGRAL	Likelihood = -4.04	Transmembrane	422 - 438 (420 - 445)
	INTEGRAL	Likelihood = -3.72	Transmembrane	365 - 381 (364 - 385)
	INTEGRAL	Likelihood = -3.40	Transmembrane	184 - 200 (182 - 202)
	INTEGRAL	Likelihood = -3.08	Transmembrane	346 - 362 (345 - 363)
	INTEGRAL	Likelihood = -1.38	Transmembrane	260 - 276 (260 - 276)

20 ----- Final Results -----

bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 A related GBS nucleic acid sequence <SEQ ID 9865> which encodes amino acid sequence <SEQ ID 9866> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA53697 GB:X76083 uracil permease [Bacillus caldolyticus]

Identities = 208/416 (50%), Positives = 291/416 (69%), Gaps = 11/416 (2%)

30	Query: 32	LLDIDEKPELFQGLLSFQHFVAMFGATILVPLILGMPVSVALFASGCGTLIYQVATKFK 91
		+LDI ++P + Q + LS QH+FAMFGATILVP ++G+ S+AL SG GTL + + TK++
	Sbjct: 5	VLDIQDRPTVGQWITLSLQHLFAMFGATILVPYLVGLDPSIALLTSGGLTFLAFLITKWQ 64
35	Query: 92	VPVYLGSSSFAYITAMALAMQMHGDISAAQTGILFVGLIYVVVATVIKFGVNSWVDKILP 151
		VP YLGSSSFAYI + A + G AA G GL+Y VVA +IK G WV K+LP
	Sbjct: 65	VPAYLGSSSFAYIAPIIAA--KTAGGPGAAMIGSFLAGLVYGVVALIHKAGYRWVMKLLP 122
40	Query: 152	PIIIGPMIIVIGLGLANSVITNA--GFVAKGDWRKMLVAVVTFLIAAFINTKGKGFIIKII 209
		P+++GP+IIVIGLGLA +AV A G K VA+VT + +G + +I
	Sbjct: 123	PVVVGPIIIVIGLGLAGTAVGMAMNGPDGKYSLHFSVALVTLAATIVCSVLARGMLSLI 182
45	Query: 210	PFLFAIIGGYILSIILGLVDLSPVEKAAWFELPKFYLFPKFTGLFHSYKLYFGPEMLAIL- 268
		P L I+ GY+ ++ +GLVDLS V A WFE P F +PF Y + E++ ++
	Sbjct: 183	PVLVGIVVGYLYALAVGLVDLSKVAAAKWFEWPDFLIPFA-----DYPVRVTWEIVMLMV 237
50	Query: 269	PISIVTIAENIGDHTVLGQICGRNFLKKPGLNRLIGDGLATAFSALIGGPAETTYGENT 328
		P++IVT++E+IG VL ++ GR+ ++KPGL+R ++GDG AT SAL+GGP +TTYGEN
	Sbjct: 238	PVAIVTLSEHIGHQLVLSKVVRDLIQKPGHLHRSILGDGTATMISALLGGPPKTTYGENI 297
55	Query: 329	GVIGMTRIASVTIRNAAFIAIAFSFFGKFTALISTIPSAVLGGMAILLYGVIASNGLKV 388
		GV+ +TR+ SV V+ AA IAIAF F GK TALIS+IP+ V+GG++ILL+G+IAS+GL++
	Sbjct: 298	GVLAITRVYSVYVLAGAAVIAIAFGFVGKITALISSIPTPVMGGVSILLFGTIASSGLRM 357
	Query: 389	LIENRVNFAEVRNLIIASSMLVLGLGGAVLDLG-ALTLSGTALSAIVGIILNLILP 443
		LI++RV+F + RNL+IAS +LV+G+GGAVL + + ++G ALSAIVG++LNLILP
	Sbjct: 358	LIDSRVDFGQTRNLVIASVILVIGIGGAVLKISDSFQITGMALSIVGVLLNLILP 413

-1965-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5431> which encodes the amino acid sequence <SEQ ID 5432>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

```

5    INTEGRAL    Likelihood = -10.72    Transmembrane    177 - 193 ( 171 - 206)
      INTEGRAL    Likelihood = -8.55     Transmembrane    313 - 329 ( 304 - 339)
      INTEGRAL    Likelihood = -8.17     Transmembrane    154 - 170 ( 152 - 175)
      INTEGRAL    Likelihood = -7.91     Transmembrane    376 - 392 ( 374 - 395)
      INTEGRAL    Likelihood = -7.48     Transmembrane    25 - 41 ( 22 - 43)
10   INTEGRAL    Likelihood = -5.84     Transmembrane    120 - 136 ( 116 - 142)
      INTEGRAL    Likelihood = -4.99     Transmembrane    96 - 112 ( 90 - 117)
      INTEGRAL    Likelihood = -3.29     Transmembrane    339 - 355 ( 338 - 360)
      INTEGRAL    Likelihood = -1.91     Transmembrane    396 - 412 ( 396 - 413)

15   ----- Final Results -----
      bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

20 The protein has homology with the following sequences in the databases:

>GP:CAB89870 GB:AJ132624 uracil transporter [Lactococcus lactis]

Identities = 294/421 (69%), Positives = 359/421 (84%), Gaps = 5/421 (1%)

```

25   Query: 3    DVIYDVEEVPKAGMLVGLSFQHLFAMFGATVLPILVGDIDPSVALLSSGLGTLAHL SVTK 62
      D+I V+E P A    GLSFQHLFAMFG+TVLPILVGI+P++ALLSSGLGTLAH+SVTK
      Sbjct: 5    DIILKVDEKPAASQWFGLSFQHLFAMFGSTVLPILVGINPAIALLSSGLGTLAHMSVTK 64

      Query: 63   FKIPAYMGSSFAYIAAMQLLTKTNGIGAVAQGAMTGGLVYLIVLIVKAIGNDWIDN ILP 122
      FK+PAYMGSSFAYI AM LLMK G+ A+AQGAMTGGLVYLIVLIVK G WID +LP
30   Sbjct: 65   FKVPAYMGSSFAYIGAMTLLMKNGGMPAIAQGAMTGGLVYLIVLIVKFAGKGWIDKVLP 124

      Query: 123  PIVVGPIVMVIGLSLASTAVNDVMLKN----GNYNLTYLVLIGLVTLSSVIFFN IYKGKIV 178
      PIVVGPIVMVIGLSLA TA+ND M + Y+L Y++I L+T+LS++ ++IYKGK +
35   Sbjct: 125  PIVVGPIVMVIGLSLAPTAINDAMYTDVANLKGYSLAYIIIALITVLSIVVYSIYKGKFL 184

      Query: 179  AIVPLLLGLLVGYVVALLVGLTQGEIVDFTNVAQAKWFSIPSVEIPFLTYGVKFYPSAI 238
      ++VP+LLG++ GYV A+++G +TG IV FT ++QAKW ++P +EIPF +Y FYPSAI
      Sbjct: 185  SVVPILLGIITGYVAAMIIGKITGMNIVSFTGISQAKWLTLPPEIIPFASYKWFYPSAI 244

40   Query: 239  LTMAPIAFVTMTEHFGHIMVLNSLTGRDYFKDPLEKTLTGDFGAQIIAGFLGAPPVTSY 298
      LTMAPIAFVTMTEHFGHIMVLNSLTG+DYFK+PGLKTLTG DG AQIIAGF+GAPPVTSY
      Sbjct: 245  LTMAPIAFVTMTEHFGHIMVLNSLTGKDYFKEPGLEKTLTG DGLAQIIAGF IGAPPVTSY 304

      Query: 299  GENIGVMALNKIFSVYVIAGAAVIAALLSFIGKVSALIQSIPTPVIGGISVALFGVIASS 358
      GENIGVMA+ KI S+YVIAGAAV+A ++SF+GK++AL+QSIP PVIGG S+ALFGVIA+S
45   Sbjct: 305  GENIGVMAITKIHSTYVIAGAAVLAIVVSFVGKITALLQSIPAPVIGGASIALFGVIAAS 364

      Query: 359  GLKILIESKVDMDNKKNLLIASVILVSGIGGLMLQV-NGLQISGVAFSTLLGIILYQVLPE 418
      GLKIL+E+KVD D K+NLLI+SV+LV GIGG+++ + LQIS VA +T+LGI+L VLP+
50   Sbjct: 365  GLKILVENKVDFDIKRNLLISSVVLVIGIGGMIINITQNLQISSVAIATILGIVLNLVLPK 425

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 186/425 (43%), Positives = 282/425 (65%), Gaps = 17/425 (4%)

```

55   Query: 30    NLLLDIDEKPELFQGLLSFQHV FAMFGATILVPLILGMPVSVL FASGCGTLIYQVATK 89
      +++ D++E P+    + LSFQHV FAMFGAT+LVP+++G+ SVAL +SG GTL + TK
      Sbjct: 3    DVIYDVEEVPKAGMLVGLSFQHLFAMFGATVLPILVGDIDPSVALLSSGLGTLAHL SVTK 62

      Query: 90    FKVPVYLGSSFAYITAMALAMQM HGDISA AQTGILFVGLIYVVVATVIK FVGN SWVDKI 149
      FK+P Y+GSSFAYI AM L MK I A G + GL+Y++VA ++K +GN W+D I
60   Sbjct: 63    FKIPAYMGSSFAYIAAMQLLTKT--NGIGAVAQGAMTGGLVYLIVLIVKAIGNDWIDNI 120

      Query: 150   LPPIIIGPMIIVIGLGLANS AVTNAGFVAKGDWRK--MLVAVVTFLIAAFINTKGKGF I K 207
      LPPI++GP+++VIGL LA++AV + + G++ +++ +VT L F N GKG +

```

-1966-

Sbjct: 121 LPPIVVGPIVMVIGLSLASTAVNDV-MLKNGNYNLTYLVIGLVTLTLLSVIFFNIYGKGIVA 179

Query: 208 IIPFLFAIIGGYILSIILG-----LVDLSPVEKAAWFELPKFYLPPFKTGLFHSYKLYFG 261
I+P L ++ GY+++++G +VD + V +A WF +P +PF T Y + F

5 Sbjct: 180 IVPLLLGLLVGYVVALLVGVLTGQEIVDFTINVAQAKWFSIPSVEIPFLT-----YGVKFY 234

Query: 262 PE-MLAILPISIVTIAENIGDHTVLGQICGRNFLKKPGLNRLLIGDGLATAFSALIGGPA 320
P +L + PI+ VT+ E+ G VL + R++ K PGL + L GDG A + +G P

10 Sbjct: 235 PSAILTMPIAFVTMTTEHFGHIMVLNSLTNRDYFKDPGLEKTLTGDFGAQIIAGFLGAPP 294

Query: 321 ETTYGENTGVIGMTRIASVTVIRNAAFIAIAFSFFGKFTALISTIPSAVLGGMAILLYGV 380
T+YGEN GV+ + +I SV VI AA IA SF GK +ALI +IP+ V+GG+++ L+GV

Sbjct: 295 VTSYGENIGVMALNKIFSIVYIAGAAVIAALLSFIGKVSALIQSIPTPVIGGISVALFGV 354

15 Query: 381 IASNGLKVLINRVNFAEVRNLIIASSMLVLGLGAVLDLGLTSLGTAISAIVGIILNL 440
IAS+GLK+LIE++V+ +NL+IAS +LV G+GG +L + L +SG A S ++GIIL

Sbjct: 355 IASSGLKILIESKVDMDNKKNLLIASVILVSGIGGLMLQVNLQISGVAFSTLLGIILYQ 414

Query: 441 ILPKE 445
+LP++

20 Sbjct: 415 VLPEK 419

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1749

A DNA sequence (GBSx1856) was identified in *S.agalactiae* <SEQ ID 5433> which encodes the amino acid sequence <SEQ ID 5434>. Analysis of this protein sequence reveals the following:

Possible site: 20
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3863 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 1750

A DNA sequence (GBSx1857) was identified in *S.agalactiae* <SEQ ID 5435> which encodes the amino acid sequence <SEQ ID 5436>. This protein is predicted to be sodium/alanine symporter. Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have no N-terminal signal sequence

45

INTEGRAL	Likelihood = -10.88	Transmembrane	191 - 207 (184 - 214)
INTEGRAL	Likelihood = -8.97	Transmembrane	151 - 167 (148 - 171)
INTEGRAL	Likelihood = -8.39	Transmembrane	217 - 233 (216 - 238)
INTEGRAL	Likelihood = -6.74	Transmembrane	312 - 328 (310 - 333)
50 INTEGRAL	Likelihood = -6.26	Transmembrane	357 - 373 (349 - 376)
INTEGRAL	Likelihood = -5.10	Transmembrane	424 - 440 (422 - 441)
INTEGRAL	Likelihood = -5.04	Transmembrane	396 - 412 (390 - 417)
INTEGRAL	Likelihood = -0.37	Transmembrane	25 - 41 (25 - 41)

55 ----- Final Results -----

-1967-

bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9867> which encodes amino acid sequence <SEQ ID 9868> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC22541 GB:U32770 amino acid carrier protein, putative
 [Haemophilus influenzae Rd]
 10 Identities = 255/443 (57%), Positives = 333/443 (74%), Gaps = 4/443 (0%)

Query: 11 TLFTTHINSFVWGPPLALLVGTGIYLSFRLGFIQLRQLSRAFKLIFREDNG-QGDISSYA 69
 ++ + I+SF+WG PLL LL GTG+YL+ RLGFIQ+R L RA +F++D G +GD+SS+A
 15 Sbjct: 5 SILSAIDSFIVGAPLLILLSGTGLYLTLRLGFIQIRYLPRALGYLFKKDKGGKGDVSSFA 64

Query: 70 ALATALAATVGTGNIVGVATAIKSGGPGALFWMWVAFFGMATKYAEGLLAIKYRTKDTN 129
 AL TALAAT+GTGNIVGVATA+++GGPGA+FMMW+ A GMATKYAE LLA+KYR +D N
 Sbjct: 65 ALCTALAATIGTGNIVGVATAVQAGGPGAIFMMWLVALGGMATKYAECLLAVKYRVRDKN 124

20 Query: 130 GEISGGPMYYIINGMGQKWKPLAVFFSAAGILVALLGIGTFTQVNAIASSLEHTFKISTR 189
 G ++GGPMYYI G+G +W LA F+ G++VA GIGTF QVNAI +++ TF I
 Sbjct: 125 GFMAGGPMYYIERGLGIRW--LAKLFALFGVMVAFFGIGTFFQVNAITHAMQDTFNIPVL 182

25 Query: 190 FTSLILAVIVLFIIFGGIKSISKVSEKIVPFMAISYILATLIIIAVNYNKIPHTFQLIFS 249
 T++I+ ++V II GG+K I+ S IVPFMAI Y+ +L+II +N K+P LI
 Sbjct: 183 VTAIIVTLLVGLIILGGVKRIATASSVIVPFMAILYVTTSLVIIILLNIEKVPDAILLIID 242

30 Query: 250 GAFSGTAAIGGFSGAIVKEAIQKGIARGVFSNESGLGSAPIAAAAAKTKEPVEQGLISMT 309
 AF AA+GG G V +AIQ G+ARG+FSNESGLGSAPIAAAAA+T+EPV QGLISMT
 Sbjct: 243 SAFDPQAAALGGAVGLTVMKAIQSGVARGIFS NESGLGSAPIAAAAAQTRFPVRQGLISMT 302

35 Query: 310 GTFIDTIVICTLTGIALVLTGWLEFDLQGAPLTQASFNTVFG-SLGSFALTFCFLVLF 368
 GTF+DTI++CT+TGI +++TG W +L GA +T +F G S+G+ +T L+ FAF
 Sbjct: 303 GTFDLTIIVCTMTGIVLVLTGAWNNPELAGATVTNYAFAQGLGTSIGATIVTVGLLFFAF 362

40 Query: 369 TTILGWSYYGERCFEYLFGTGKFINAYRIIFVIMVGLGGFLQLDLIWIADIIVNGLMALPN 428
 TTILGW YYGERCF YL G + + YR+ ++++VGLG FL L+LIW+IADIIVNGLMA PN
 Sbjct: 363 TTILGWYCYGERCFVYLVGIRGVKLYRLAYIMLVGLGAFLHLNLIWIADIIVNGLMAFPN 422

40 Query: 429 LIALALLSPIIVKETQKYFSETK 451
 LIAL+ L +I++ET+ YF K
 Sbjct: 423 LIALIGLRKVIEETKDYFQRLK 445

- 45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5437> which encodes the amino acid sequence <SEQ ID 5438>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -11.36	Transmembrane	183 - 199 (175 - 206)
INTEGRAL	Likelihood = -7.80	Transmembrane	143 - 159 (140 - 163)
INTEGRAL	Likelihood = -7.11	Transmembrane	209 - 225 (208 - 229)
INTEGRAL	Likelihood = -5.95	Transmembrane	416 - 432 (413 - 434)
INTEGRAL	Likelihood = -5.15	Transmembrane	304 - 320 (302 - 324)
INTEGRAL	Likelihood = -4.46	Transmembrane	387 - 403 (382 - 408)
INTEGRAL	Likelihood = -3.35	Transmembrane	348 - 364 (345 - 366)
INTEGRAL	Likelihood = -1.17	Transmembrane	11 - 27 (10 - 28)

50

----- Final Results -----

bacterial membrane --- Certainty=0.5543(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60

The protein has homology with the following sequences in the databases:

-1968-

>GP:AAF94579 GB:AE004221 sodium/alanine symporter [Vibrio cholerae]
Identities = 261/441 (59%), Positives = 328/441 (74%), Gaps = 7/441 (1%)

Query: 3 ALVKLIDNLVWGPPDLLLVGTGIYLTSHLGLIQILKLPRAFKLIFSDDEG---HGDISS 59
+ ++ +D+LVWGPPDLLLVGTG+Y T LGL+Q +LP A ++F ++ GD+SS
Sbjct: 6 SFLQTVDSLWVGPPDLLLVGTGVYFTFRLGLLQFRRRLPTALAMVFGREKSSDKQGDVSS 65

Query: 60 FAALATALAATVGTGNIVGVATAIKSGGPGALFWMWVAFFFGMATKYAEGVLAIKYRTKD 119
FAAL TAL+AT+GTGNIVGVATAIK GGPALFWMW+AA FGMATKYAE +LA+KYR D
Sbjct: 66 FAALCTALSATIGTGNIVGVATAIKLGGPGALFWMWLALFGMATKYAECLLAVKYRQID 125

Query: 120 ANGHISGGPMYYIVNGMGTKWKPLAVLFAGSGILVALFGIGTFAQVNSITSSLGHSFGLS 179
G + GGPMYY+ +G+ +K LAVLFA + VA FGIGTF QVN+I + SFG+
Sbjct: 126 DKGQMVGGPMYYLRDGVSSK--TLAVLFAVFAVGACFGIGTFQVNAILDATQISFGVP 183

Query: 180 PQMVSIVLAIFVAAIIFGGIHSISKVAEKVVPFMAIFYILSSLAVIFSHYQQLLPVIRLV 239
+ ++VL + VA + GGI SI+KVA KVVP MA+FYI++ L+VI ++ +L + LV
Sbjct: 184 REASAVVLTVLVAIVTIGGIQSIKAVAGKVVPAMALFYIIACLSVIVTNADKLADAVELV 243

Query: 240 FQSFTPTAAIGGFAGSLMKDAIQKGIARGVFSNESGLRSAPIAAAAAKTNEPVEQGLIS 299
SAFT TAA GGF G+ + AIQ GIARGVFSNESGL SAP+AAAAAKT+ VEQGLIS
Sbjct: 244 LVSFTSTAATGGFLGASIMLAIQSGIARGVFSNESGLSAPMAAAAKTDCSVEQGLIS 303

Query: 300 MTGTFDITIIICTLTGLSILVTGQWTGQLEGAPLTQSAFATVFG--NLGTFGLTFSLVLF 357
MTGTF DTIIICT+TGL++++TG W L GA +T AFAT +G ++ L+ F
Sbjct: 304 MTGTFDITIIICTMTGLALILTGAWQSDLSGAAMTTYAFATGLNAQTIGPMLVSIQGLMEF 363

Query: 358 AFTTILGWSYYGERCFEFLFGITHLTYFRIVFILMVGLGGFLKLELIWVLADIVNGLMAL 417
AFTTILGW+YYGERC FLFG + ++IVFI ++ G FL L+LIW++ADIVNGLMA+
Sbjct: 364 AFTTILGWNYYGERCMVFLFGTKAVLPYKIVFIGLIASGAFLHLDLIWIADIVNGLMAI 423

Query: 418 PNLIALLALSPVILETKHYF 438
PNLI L+AL VV+ ETK YF
Sbjct: 424 PNLIGLVALRHVVVEETQYF 444

An alignment of the GAS and GBS proteins is shown below.

Identities = 323/439 (73%), Positives = 380/439 (85%), Gaps = 1/439 (0%)

Query: 9 MLTLFTHINSFVWGPPLALLLVGTGIYLSFRLGFIQLRQLSRAFKLIFREDNGQGDISSY 68
M+ L I++ VWGPPLL LLVGTGIYL+ LG IQ+ +L RAFKLIF +D G GDISS+
Sbjct: 1 MIALVKLIDNLVWGPPDLLLVGTGIYLTSHLGLIQILKLPRAFKLIFSDDEGHGDISSF 60

Query: 69 AALATALAATVGTGNIVGVATAIKSGGPGALFWMWVAFFFGMATKYAEGLLAIKYRTKDT 128
AALATALAATVGTGNIVGVATAIKSGGPGALFWMWVAFFFGMATKYAEG+LAIKYRTKD
Sbjct: 61 AALATALAATVGTGNIVGVATAIKSGGPGALFWMWVAFFFGMATKYAEGVLAIKYRTKDA 120

Query: 129 NGEISGGPMYYIINGMGQWKPLAVFFSAAGILVALLGIGTFTQVNAIASSLEHTFKIST 188
NG ISGGPMYYI+NGMG KWKPLAV F+ +GILVAL GIGTF QVN+I SSL H+F +S
Sbjct: 121 NGHISGGPMYYIVNGMGTKWKPLAVLFAGSGILVALFGIGTFAQVNSITSSLGHSFGLSP 180

Query: 189 RFTSLILAVIVLFIIFGGIKSISKVSEKIVPFMAISYILATLIIIVNYPKIPHTFQLIF 248
+ S++LA+ V IIFGGI SISKV+EK+VPFMAI YIL++L +I +Y ++ +L+F
Sbjct: 181 QMVSIVLAIFVAAIIFGGIHSISKVAEKVVPFMAIFYILSSLAVIFSHYQQLLPVIRLVF 240

Query: 249 SGAFSGTAAIGGFSGAIVKEAIQKGIARGVFSNESGLSAPIAAAAAKTKEPVEQGLISM 308
AF+ TAAIGGF+G+++K+AIQKGIARGVFSNESGL SAPIAAAAAKT EPVEQGLISM
Sbjct: 241 QSAFTPTAAIGGFAGSLMKDAIQKGIARGVFSNESGLRSAPIAAAAAKTNEPVEQGLISM 300

Query: 309 TGTFTDITIVICTLTGIAILVTGKWLEFDLQCAPLTQASFNTVFGSLGSFALTFCILVLF 368
TGTFTDITI+ICTLTG++ILVTG+W L+GAPLTQ++F TVFG+LG+F LTF LVLFAF
Sbjct: 301 TGTFTDITIIICTLTGLSILVTGQWTG-QLEGAPLTQSAFATVFGNLGTFGLTFSLVLF 359

Query: 369 TTILGWSYYGERCFEFLFGTKFINAYRIIFVIMVGLGGFLQLDLIWIADIVNGLMALPN 428
TTILGWSYYGERCFE+LFG + +RI+F++MVGLGGFL+L+LIWV+ADIVNGLMALPN
Sbjct: 360 TTILGWSYYGERCFEFLFGITHLTYFRIVFILMVGLGGFLKLELIWVLADIVNGLMALPN 419

Query: 429 LIALLALSPIIVKETQKYF 447

-1969-

LIALLALSP+++ ET+ YF
 Sbjct: 420 LIALLALSPVVILETKHYF 438

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1751

A DNA sequence (GBSx1858) was identified in *S.agalactiae* <SEQ ID 5439> which encodes the amino acid sequence <SEQ ID 5440>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -6.16	Transmembrane	85 - 101 (80 - 108)
INTEGRAL	Likelihood = -5.36	Transmembrane	118 - 134 (115 - 137)
INTEGRAL	Likelihood = -2.81	Transmembrane	177 - 193 (177 - 193)
INTEGRAL	Likelihood = -0.48	Transmembrane	49 - 65 (49 - 65)

----- Final Results -----

bacterial membrane	---	Certainty=0.3463(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12451 GB:Z99107 alternate gene name: ydxT~similar to cation
 efflux system membrane protein [Bacillus subtilis]
 Identities = 118/282 (41%), Positives = 181/282 (63%)

Query: 6 ENLQLAKRGPIISIIAYITLAVAKLAAGYWF DATSLVADGFNNLS DILGNVALLIGLHLA 65
 + L+ + G ++SI AY+ L+ KL GY F + +L ADG NN +DI+ +VA+LIGL ++
 Sbjct: 5 DELKKGESGALVSIAAYLVLSAIIIGYLFHSEALTADGLNNTTDIIASVAVLIGLRIS 64

Query: 66 SQPADSNHRFGHWKIEDLASLITSFIMFVVG IQVFIQT VTKIINNTDTNIDPLGAIVGAI 125
 +P D +H +GH++ E +ASLI SFIM VVG+QV I + D + A A
 Sbjct: 65 QKPPDEDHPYGHFRAETIASLIASFIMMVVGLQVLF SAGESIFS AKQETPD MIAAWTAAG 124

Query: 126 SALVMLGVYFYFNKQLSQRVKSSALVAASKDNLSDAVTSIGT SIAIIAASINFPPIIDRLAA 185
 A++ML VY YNK+L+++VKS AL+AA+ DN SDA SIGT I I+AA + ID + A
 Sbjct: 125 GAVLMLIVRYRNKRLAKVKVKSQALLAAADNKSDAFV SIGTFIGIVAAQFHLAWIDT VTA 184

Query: 186 IIITYFILKTAYDIFIESAFSLSDGFD DQYQYKAILTI PKISAVKSQRGR TYGSNIY 245
 +I I KTA+DIF ES+ SL+DGFD + Y++ I I +S +K + R GS ++
 Sbjct: 185 FVIGLLICKTAWDIFKESHSLTDGFDIKDISAYKQTIEKISGVSR LKDIKARYLGSTVH 244

Query: 246 LDIVLEMNPDLSVFESHAI TERVEKLLSDKFSVYDIDIHVEP 287
 +D+V+E++ DL++ ESH I +E+ + ++ ++ +H+EP
 Sbjct: 245 VDVVVEVSADLNITESHDIANEIERRMKEEHAIDYSHVHMEP 286

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5441> which encodes the amino acid sequence <SEQ ID 5442>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -8.01	Transmembrane	121 - 137 (114 - 139)
INTEGRAL	Likelihood = -5.41	Transmembrane	86 - 102 (84 - 109)
INTEGRAL	Likelihood = -5.04	Transmembrane	178 - 194 (176 - 197)
INTEGRAL	Likelihood = -0.69	Transmembrane	50 - 66 (50 - 66)
INTEGRAL	Likelihood = -0.64	Transmembrane	158 - 174 (158 - 174)

----- Final Results -----

bacterial membrane	---	Certainty=0.4206(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

-1970-

The protein has homology with the following sequences in the databases:

>GP:CAB12451 GB:Z99107 alternate gene name: ydxT~similar to cation
efflux system membrane protein [Bacillus subtilis]
Identities = 127/280 (45%), Positives = 187/280 (66%)

5 Query: 9 LKLARKGPVSIIVYLSLSVAKLLAGYLLNASSLIADGFNNLSDIVGNVALLIGLHLASQ 68
LK G +VSI YL LS KL+ GYL ++ +L ADG NN +DI+ +VA+LIGL ++ +
Sbjct: 7 LKKGESGALVSIAYLVLSAIIKLIIGYLFHSEALTADGLNNTTDIIASVAVLIGLRISQK 66

10 Query: 69 PADANHKFGHWKIEDLSSLVTSFIMFLVGFQVLIHTIKSIFSGQQVDIDPLGAIVGIVSA 128
P D +H +GH++ E ++SL+ SFIM +VG QVL +SIFS +Q D + A A
Sbjct: 67 PPDEHPYGFHRAETIASLIASFIMMVVGLQVLFSGESIFSQKQETPDMIAAWTAAGGA 126

15 Query: 129 FVMLGVVYFNKRLSKRVKSSALVAASKDNLADAVTSIGTSAIIAASLHLPVIDHIAAMI 188
+ML VY +NKRL+K+VKS AL+AA+ DN +DA SIGT I I+AA HL ID + A +
Sbjct: 127 VLMLIVRYNKRLLAKKVKSQLALAAADNKSDFVSIGTFIGIVAAQFHLAWIDTVTAFV 186

20 Query: 189 ITFFILKTAFDIFMESSFSLSDGFDSRHLKKYEKAILIPIKIVAVKSQRARTYGSNVYLD 248
I I KTA+DIF ESS SL+DGFD + + Y++ I +I + +K +AR GS V++D
Sbjct: 187 IGLLICKTAWDIFKESSHSLTDGFDIKDISAYKQTIEKISGVSRLLKDIKARYLGSTVHVD 246

25 Query: 249 IVLEMNPDLVYESHSITEKVEQLLSDQFSIYDIDIHVEP 288
+V+E++ DL++ ESH I ++E+ + ++ +I +H+EP
Sbjct: 247 VVVEVSADLNITESHDIANEIERRMKEHAIDYSHVHMEP 286

An alignment of the GAS and GBS proteins is shown below.

Identities = 274/406 (67%), Positives = 340/406 (83%), Gaps = 4/406 (0%)

30 Query: 7 NLQLAKRGPIISIIAYITLAVAKLAAGYWF DATSLVADGFNNLSDIIGNVALLIGLHLAS 66
NL+LA++GPI+SII Y++L+VAKL AGY +A+SL+ADGFNNLSDI+GNVALLIGLHLAS
Sbjct: 8 NLKLARKGPVSIIVYLSLSVAKLLAGYLLNASSLIADGFNNLSDIVGNVALLIGLHLAS 67

35 Query: 67 QPADSNHRFGHWKIEDLASLITSFIMFVVGIVFIQTIVTKIINNTDTNIDPLGAIVGAIS 126
QPAD+NH+FGHWKIEDL+SL+TSFIMF+VG QV I T+ I + +IDPLGAIVG +S
Sbjct: 68 QPADANHKFGHWKIEDLSSLVTSFIMFLVGFQVLIHTIKSIFSGQQVDIDPLGAIVGIVS 127

40 Query: 127 ALVMLGVYFYFNKQLSQRVKSSALVAASKDNLSDAVTSIGTSAIIAASLNFFIIDRLAAI 186
A VMLGVY +NK+LS+RVKSSALVAASKDNL+DAVTSIGTSAIIAASL+ P+ID +AA+
Sbjct: 128 AFVMLGVYVFNKRLSKRVKSSALVAASKDNLADAVTSIGTSAIIAASLHLPVIDHIAAM 187

45 Query: 187 IITYFILKTAYDIFIESAFSLSDGFDDYQLKQYEKAILTIPKISAVKSQRGRTYGSNIYL 246
IIT+FILKTA+DIF+ES+FSLSDGFD LK+YEKAIL IPKI AVKSQR RTYGSN+YL
Sbjct: 188 IITFFILKTAFDIFMESSFSLSDGFDSRHLKKYEKAILIPIKIVAVKSQRARTYGSNVYL 247

50 Query: 247 DIVLEMNPDLVSVFESHATERVEKLLSDKFSVYDIDIHVEPASIPEDIIDNVYQKLYKN 306
DIVLEMNPDLVSV+ESH+ITE+VE+LLSD+FS+YDIDIHVEPA IPE+EIDNV +KLY+
Sbjct: 248 DIVLEMNPDLVYESHSITEKVEQLLSDQFSIYDIDIHVEPAMIPPEEIDNVAKKLYRY 307

55 Query: 307 EKILAKIPGYETFISPDFYMINKEGNIITSDMLTNATNHSLSANFKYFNVKSISQKTKL 366
EK+IL+K+P Y+ +I+ F +I+ G + + N + SNF +F ++SISQK L
Sbjct: 308 EKLILSKVPDYDHYIAKSFQLIDANGQTVNVEQFLNQEIY-YPSNFNHFQIESISQKTKL 366

Query: 367 VSYELEGKRHTSIWRRNEKWFLIYHQIT--AKSSPYKTRRYQITSL 410
V+Y+L G + TSIWRR+E W L++HQIT AK + T Y+I +
Sbjct: 367 VTYQLNGNQRTSIWRRHESWSLLFHQITPIAKKQLHHT-HYRIVKM 411

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1752

60 A DNA sequence (GBSx1859) was identified in *S.agalactiae* <SEQ ID 5443> which encodes the amino acid sequence <SEQ ID 5444>. Analysis of this protein sequence reveals the following:

-1971-

Possible site: 55

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -8.12 Transmembrane 171 - 187 (161 - 194)
 INTEGRAL Likelihood = -6.32 Transmembrane 118 - 134 (113 - 138)
 INTEGRAL Likelihood = -5.89 Transmembrane 59 - 75 (53 - 77)
 INTEGRAL Likelihood = -5.52 Transmembrane 231 - 247 (226 - 252)
 INTEGRAL Likelihood = -3.24 Transmembrane 86 - 102 (84 - 103)
 INTEGRAL Likelihood = -0.32 Transmembrane 31 - 47 (31 - 47)

----- Final Results -----

bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9869> which encodes amino acid sequence <SEQ ID 9870> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14850 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 80/226 (35%), Positives = 136/226 (59%), Gaps = 1/226 (0%)

Query: 27 TNNPIFGIMLTWYAYIGIRIFRKYPSPAT-TPLLLATILLIAFLKLTHISYKDYNGGS 85
 T +P FGI++++ A+ IG +F+K TPL +A +L IAFLK+ SY DY NGG
 Sbjct: 4 TMSPFYGVIVSLAAPGIGTFLFKKTGFFLFTPLFVAMVLGIAFLKIGGFSYADYNNNGE 63

Query: 86 FLTMLITPSTVVLAIPLYRTFHLMKHHIKSISISIIILASVINTVFTAIVAKFFGMKYFLA 145
 + + P+T+ AIPLY+ +K + I SII S+ + ++AK + +
 Sbjct: 64 IIKFFLEPATIAFAIPLYKQDKLKKYWWQIMASIIAGSICSVTIVYLLAKGIHLDSAVM 123

Query: 146 ISLFPKSVTTAMAVGITSKAGGLATITLVVVVITGILTSVLGPIFLKLLRIEDPVAIGLA 205
 S+ P++ TTA+A+ ++ GG++ IT V+ ++ LG +FLK+ ++++P++ GLA
 Sbjct: 124 KSMPLQAATTAIALPLSKGIGGISDITAFVIFNAVIVYALGALFLKVKVKNPISKGLA 183

Query: 206 LGGTGHAIGTGQALKYGOVQCAMAGLAIGITGICYVIVSPLVAGLI 251
 LG +GHA+G ++ G+V+ AMA +A+ + G+ V+V P+ LI

Sbjct: 184 LGTSGHALGVAVGIEMGEVEAAMASIAVVVGVVTVLVIPVFVQLI 229

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8893> and protein <SEQ ID 8894> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0

SRCFLG: 0

McG: Length of UR: 22

Peak Value of UR: 2.57

Net Charge of CR: 0

McG: Discrim Score: 6.51

GvH: Signal Score (-7.5): -5.91

Possible site: 33

>>> Seems to have an uncleavable N-term signal seq

Amino Acid Composition: calculated from 1

ALOM program count: 6 value: -8.12 threshold: 0.0

INTEGRAL Likelihood = -8.12 Transmembrane 149 - 165 (139 - 172)
 INTEGRAL Likelihood = -6.32 Transmembrane 96 - 112 (91 - 116)
 INTEGRAL Likelihood = -5.89 Transmembrane 37 - 53 (31 - 55)
 INTEGRAL Likelihood = -5.52 Transmembrane 209 - 225 (204 - 230)
 INTEGRAL Likelihood = -3.24 Transmembrane 64 - 80 (62 - 81)
 INTEGRAL Likelihood = -0.32 Transmembrane 9 - 25 (9 - 25)
 PERIPHERAL Likelihood = 1.06 121

modified ALOM score: 2.12

icm1 HYPID: 7 CFP: 0.425

*** Reasoning Step: 3

-1972-

----- Final Results -----

bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF01066(325 - 999 of 1305)

EGAD|107753|BS2884(4 - 229 of 231) hypothetical protein {Bacillus subtilis} OMNI|NT01BS3363
 LrgB GP|1770004|emb|CAA99613.1||Z75208 hypothetical protein {Bacillus subtilis}
 GP|2635355|emb|CAB14850.1||Z99118 similar to hypothetical proteins {Bacillus subtilis}
 PIR|D69983|D69983 conserved hypothetical protein ysbB - Bacillus subtilis

%Match = 17.2

%Identity = 35.4 %Similarity = 62.4

Matches = 80 Mismatches = 84 Conservative Sub.s = 61

192 222 252 282 312 342 372 402
 WSTFKT*SPIFLG*LSLS*ERYFSIF*LLDWYPNGSKRDMKEIIQKLEVKMATLTNNPIFGIMLTVWAYIIGIRIFRKYP
 | : | |||::: | : | : | :
 MESTMSPYFGIVVSLAAGIGITFLFKKTK
 10 20

429 459 489 519 549 579 609 639
 SPAT-TPLLATILLIAFLKLTHISYKDYNGGSFLTMLITPSTVVLAIPLYRTFHLMKHHIKSISISIIASVINTVFT
 ||::: | : ||||: || || || : :: |:| : ||||: : | : | || | : :
 GFFLFTPLFVAMVLGIAFLKIGGFSYADYNNNGGEIIFKFFLEPATIAFAIPLYKQDKLKKYWWQIMASIIAGSICSVTIV
 40 50 60 70 80 90 100

669 699 729 759 789 819 849 879
 AIVAKFFGMKYFLAISLFPKSVTTAMAVGITSKAGGLATITLVVVITGILTSVLGPIFLKLLRIEDPVAIGLALGGTGH
 ::|| : : |:|:: |||: : : ||: || | : : || : |||:::|: |||| : ||
 YLLAKGIHLDSAVMKSMLPQAATTATALPLSKGIGGISDITAFAVIFNAVIVYALGALFLKVFKVKNPISKGLALGTSGH
 120 130 140 150 160 170 180

909 939 969 999 1029 1059 1089 1119
 AIGTQALKYGVQVQAMAGLAIGITGICVIVSPLVAGLILK*G*GK*TONNYVIFKNRI*DK*L*YR*KK*LLERLSV
 |:| : : |:|: ||| : |:| : |:| : ||
 ALGVAVGIEMGEVEAAMASIAVVVVGVVTVLVIPVFVQLIGG
 200 210 220 230

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1753

A DNA sequence (GBSx1860) was identified in *S.agalactiae* <SEQ ID 5445> which encodes the amino acid sequence <SEQ ID 5446>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA76857 GB:Y17797 hypothetical protein [Enterococcus faecalis]
 Identities = 44/194 (22%), Positives = 90/194 (45%), Gaps = 13/194 (6%)

Query: 21 TACSSSNTQQTSTSKSNVSQHKNIKADHEELRLKFNKVLGVKANNFKGGTSLAELKQLF 80
 T S ++T++ S+ K + + K D+ +L+ ++K+ +G N+ +GG++ E+K +
 Sbjct: 60 TNSSKNDTKKESSEKKSEDKSK----DNSDLKATYDKINVGDIMNSSEGGSTEDVKAIL 115
 Query: 81 GGEPNEKFDTFAGNVTLKGYRW-NVDD----ISTTIQLLNDSSIVRSISNFKFIRDANIT 135

-1973-

GEP .T ++ W NV SIT+ + + +S+S K + +T
 Sbjct: 116 -GEPASSSTTDIQISTTTLSWTNVKGGDLLASITVSFSDGKAASKSVSGLKVAHKDKVT 174

Query: 136 TKDYNLSLKNMGMSYN--KVKELLGEPDDISQAVSSDKEELQAAWISGIQSSSDSGINLTF 193
 N++ SY+ + ++ LG+P I+ + ++ W+ + D + ++F

Sbjct: 175 ADQVNNIATDGSYSEEQARKDLGDPTGITSTNINGEKNDTLIWMKNL-DGDLGATVTVSF 233

Query: 194 ENDKLTNKQQHGLK 207
 N +K GLK

Sbjct: 234 SNGNAISKSSSGLK 247

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5447> which encodes the amino acid sequence <SEQ ID 5448>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA76857 GB:Y17797 hypothetical protein [Enterococcus faecalis]
 Identities = 34/166 (20%), Positives = 74/166 (44%), Gaps = 8/166 (4%)

Query: 47 HQDKRANFEKIKLATVDSSFTGGTSLLEELISLFGEPHQHDPKTAGEVTIDAYTWQFDQ-- 104
 + D +A ++KI + + +S GG++ +E+ ++ GEP+ ++ +W +

Sbjct: 83 NSDLKATYDKINVGDIMNSSEGGSTEDVKAILGEPASSSTTDIQISTTTLSWTNVKGG 142

Query: 105 ---VTLTVNLYQNSSIVKTIISNFTFARELGLSQKEYQQLQKGMYS--EDVKILTEPDNY 159
 ++TV+ + K++S A+ ++ + + SY E +K L +P

Sbjct: 143 DLLASITVSFSDGKAASKSVSGLKVAHKDKVTADQVNNIATDGSYSEEQARKDLGDPTGI 202

Query: 160 SQASSSDHQTLQAIWVSGLKTDTSKANISLVFENNQLTEMSQVGL 205
 + + + + IW+ L D GA +++ F N S GL+

Sbjct: 203 TSTNINGEKNDTLIWMKNLDGDL-GATVTVSFSNGNAISKSSSGLK 247

An alignment of the GAS and GBS proteins is shown below.

Identities = 84/199 (42%), Positives = 126/199 (63%), Gaps = 3/199 (1%)

Query: 11 TIVCLSFGL--LTACSSSNTQQTSTSKSNVSQHKNIKADHEELRLKFNKVKLGKANNFK 68
 T++ +SF L ACS++ ++ S S + + +A H++ R F K+KL ++F

Sbjct: 8 TLLLSISFFTSFLVACSTTKDKEPQPSDSEIITPRLHQAHHQDKRANFEKIKLATVDSSFT 67

Query: 69 GGTSLAELKQLFGGEPNEKFDTPAGNVTLKGYRWNVDDISITQLLNDSSIVRSISNFKF 128
 GGTSL EL LFG EP++ AG VT+ Y W D +++T+ L +SSIV++ISNF F

Sbjct: 68 GGTSLLEELISLFG-EPHQHDPKTAGEVTIDAYTWQFDQVTLTVNLYQNSSIVKTIISNFTF 126

Query: 129 IRDANITTKDYNLSLKNMGMSYNKVKELLGEPDDISQAVSSDKEELQAAWISGIQSSSDSG 188
 R+ ++ K+Y L+ GMSY VK++L EPD+ SQA SSD + LQA W+SG+++ S

Sbjct: 127 ARELGLSQKEYQQLQKGMYSYEDVKILTEPDNYSQASSSDHQTLQAIWVSGLKTDTSKAN 186

Query: 189 INLTFENDKLTNKQQHGLK 207
 I+L FEN++LT Q GL+

Sbjct: 187 ISLVFENNQLTEMSQVGL 205

SEQ ID 5446 (GBS650) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 178 (lane 9; MW 28kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1974-

Example 1754

A DNA sequence (GBSx1861) was identified in *S.agalactiae* <SEQ ID 5449> which encodes the amino acid sequence <SEQ ID 5450>. This protein is predicted to be ribosomal protein S1 homolog; Sequence specific DNA-binding protein (r. Analysis of this protein sequence reveals the following:

```

5   Possible site: 46
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2950(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9363> which encodes amino acid sequence <SEQ ID 9364> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA97575 GB:U27517 ribosomal S1 protein [Homo sapiens]
  Identities = 156/305 (51%), Positives = 214/305 (70%), Gaps = 7/305 (2%)

Query: 1  MEARKAWDKLVGREGEVTVKGTTRAVKGGLSVEFEGLRGFIPAS MIDTRFVRNTEKFVQG 60
20  ++ARKAW+ L  EG+ V K  AV+CGL V+  G+RGF+PASM+ RFV + +F +
Sbjct: 53  LDARKAWENLSFAEGDTVDAKVINAVRGGLIVDVNGVRGFPAS MVAERFVSDLNQFKNK 112

Query: 61  EFDAKIKEVDAAENRFILSRREVVEESAAAARKEVFSNIEVGSVVTGKVARLTSFGAFID 120
25  + A++ E+D A R ILSR+ V + AA  EVFS + VG VV G VARLT FGAF+D
Sbjct: 113 DIKAQVIEIDPANARLILSRKAVAAQERAAQLAEVFSKLSVGEVVEGTVARLTDGFAFVD 172

Query: 121 LGGVDGLVHVTELSHERNVSPKSVVTVGEEVEVKVLSIDEEAGRVSLSLKATTPGPWDGV 180
30  LGGVDGLVHV+E+SH+R +P  V+T G++V+VK+L++D E GR+SLS+KAT GPWD
Sbjct: 173 LGGVDGLVHVSEISHDRVKNPADVLTKGDKVDVKILALDTEKGRISLSIKATQRGPWDEA 232

Query: 181 EQKLAAGDVIEGKVKRLTDFGAFVEVLPGIDGLVHISQISHKRVENPKDVL SAGQEVTVK 240
35  ++AAG V+EG VKR+ DFGAFVE+LPGI+GLVH+SQIS+KR+ENP +VL +G +V VK
Sbjct: 233 ADQIAAGSVLEGTVKRVKDFGAFVEILPGIEGLVHVSQISNKRIENPSEVLKSGDKVQVK 292

Query: 241 VLEVNSDAERVSLSMKALEERPAQAEKEKEKRQSRPRRPRRQEKRDYELPETQTGF SMA 300
40  VL++  ER+SLSMKALEE+P  + E R+  R +  Y+ + + ++
Sbjct: 293 VLDIKPAEERISLSMKALEEKP-----EREDRRGNDGSASRADIAAYK-QQDDSAATLG 345

Query: 301 DLFGD 305
40  D+FGD
Sbjct: 346 DIFGD 350

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5451> which encodes the amino acid sequence <SEQ ID 5452>. Analysis of this protein sequence reveals the following:

```

45   Possible site: 26
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
50      bacterial cytoplasm --- Certainty=0.3312(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 284/309 (91%), Positives = 296/309 (94%), Gaps = 1/309 (0%)
55
Query: 1  MEARKAWDKLVGREGEVTVKGTTRAVKGGLSVEFEGLRGFIPAS MIDTRFVRNTEKFVQG 60
      +EARKAWDKLVGREGEVTVKGTTRAVKGGLSVEFEGLRGFIPAS MIDTRFVRNTEKFVQG
Sbjct: 93  LEARKAWDKLVGREGEVTVKGTTRAVKGGLSVEFEGLRGFIPAS MIDTRFVRNTEKFVQG 152

```


-1975-

Query: 61 EFD AKI KEVDAAENR FILSRREVVEESAAAARKEVFSNIEVGSVVTGKVARLTSFGAFID 120
 EFD AKI KEVDAAENR FILSRREV+EE+A AR EVFS I G+VVTG VARLTSFGAFID
 Sbjct: 153 EFD AKI KEVDAAENR FILSRREVIEEAAKEARAEVFSKISEGAVVTGTVARLTSFGAFID 212

5 Query: 121 LGGVDGLVHVTEL SHERNVSPKSVVTVGEEVEVKVLSIDEEAGRVSLSLKATTPGPWDGV 180
 LGGVDGLVHVTEL SHERNVSPKSVV+VGEEVEVKVLSIDEEAGRVSLSLKATTPGPWDGV
 Sbjct: 213 LGGVDGLVHVTEL SHERNVSPKSVVSVGEEVEVKVLSIDEEAGRVSLSLKATTPGPWDGV 272

10 Query: 181 EQKLAAGDVIEGKVRLTDFGAFVEVLPGIDGLVHISQISHKRVENPKDVL SAGQEVTVK 240
 EQKLA GDV+EGKVRLTDFGAFVEVLPGIDGLVHISQISHKRVENPKDVL S GQEVTVK
 Sbjct: 273 EQKLAQGDVVEGKVRLTDFGAFVEVLPGIDGLVHISQISHKRVENPKDVL SVGQEVTVK 332

15 Query: 241 VLEVN SDAERVSL SMLKALEERPAQAEGE-KEEKQRSRPRRPRRQEKRDYELPETQTGFMS 299
 VLEVN+ ERVSL S+KALEERPAQAE G+ KEEKQRSRPRR+R+ +RDYELPETQTGFMS
 Sbjct: 333 VLEVN AADERVSL SIKALEERPAQAE GDNKEEKQRSRPRRPKRESRDYELPETQTGFMS 392

20 Query: 300 ADLFGDIEL 308
 ADLFGDIEL
 Sbjct: 393 ADLFGDIEL 401

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1755

A DNA sequence (GBSx1862) was identified in *S.agalactiae* <SEQ ID 5453> which encodes the amino acid sequence <SEQ ID 5454>. This protein is predicted to be dihydroorotate dehydrogenase a (pyrD). Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1708(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB51330 GB:AJ131985 dihydroorotate dehydrogenase [Streptococcus pneumoniae]
 Identities = 227/310 (73%), Positives = 268/310 (86%)

Query: 1 MVSLKTEIAGFSFDNCLMNAAGIYCM TKEELLAIENSEAGSFVTKTGTLEAREGNPQPRY 60
 MVS KT+IAGF FDNCLMNAAG+ CMT EEL ++NS AG+FVTKT TL+ R+GNP+PRY
 Sbjct: 1 MVSTK TQIAGFEFDNCLMNAAGVACMTIEELEEVKNSAAGTFVTKTATLDFRQGNPEPRY 60

Query: 61 ADTDWGSINSMGLPNKGIDYYLDFVTELQDQDNSKNHVL SLVGLSPEETHIILKKVENS 120
 D GSINSMGLPN G+DYYLD++ +LQ+++++ LSLVG+SPEETH IILKV+ S
 Sbjct: 61 QDVPLGSINSMGLPNNGLDYYLDYLLDLQEKESNRTFFLSLVGMSPEETHIILKKVQESD 120

Query: 121 YNGLIELNLSCPNVPGKPQIAYDFEMTDLILSEIFSYQKPLGIKLPYFDIVHFDQAAT 180
 + GL ELNLSCPNVPGKPQIAYDFE TD IL+E+F+Y+ KPLGIKLPYFDIV+FDQAA
 Sbjct: 121 FRGLTELNLSCPNVPGKPQIAYDFETTDRLAEVFAVFTKPLGIKLPYFDIVFDQAAA 180

Query: 181 IFNKYPLAFINCVNSIGNGLVIDDET VVIKPKNGFGGIGGDFIKPTALANVHAFYKRLNP 240
 IFNKYPL F+NCVNSIGNGL I+DE+VVI+PKNGFGGIGG++IKPTALANVHAFY+RLNP
 Sbjct: 181 IFNKYPLKFVNCVNSIGNGLYIEDES VVIRPKNGFGGIGGEYIKPTALANVHAFYQRLNP 240

Query: 241 SIKIIGTGGVKNGRDAFEHILCGASMVQIGTALQKEGPEIFQVRSRELKEIMADKGYQSL 300
 I+IIGTGGV GRDAFEHILCGASMVQ+GT L KEG F R++ ELK IM +KGY+SL
 Sbjct: 241 QIQIIGTGGVLTGRDAFEHILCGASMVQVGTTLHKEGVSAFDRITNELKAIMVEKGYESL 300

Query: 301 EDFRGQLNYL 310
 EDFRG+L Y+
 Sbjct: 301 EDFRGKLR YI 310

-1976-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5455> which encodes the amino acid sequence <SEQ ID 5456>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2689(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 239/309 (77%), Positives = 262/309 (84%)

Query: 1 MVSLKTEIAGFSFDNCLMNAAGIYCMCKEELLAIENSEAGSFVTKTGTLEAREGNPQPRY 60
MVS T+I FSDNCLMNAAG+YCMCKEEL+ +E S+A SFVTKTGTLE R GNP+PRY
Sbjct: 5 MVSTATQIGHFSFDNCLMNAAGVYCMCKEELMEVEKSQAASFVTKTGTLEVRPGNPEPRY 64

Query: 61 ADTDWGSINSMLPNKGIDYYLDFVTELDQDQNSKNHVLVLVGLSPEETHIILKKVENSS 120
ADT GSINSMLPN G YYLDFV++L K H LS+VGLSP ET ILK + S
Sbjct: 65 ADTRLGSINSMLPNNGFRYYLDFVSDLAKTGQHKPHFLSVVGLSPTETETILKAIMASD 124

Query: 121 YNGLIELNLSCPNVPGKQIAYDFEMTDLILSEIFSYQKPLGIKLPFYFDIVHFDQAAT 180
Y GL+ELNLSCPNVPGKQIAYDFE TD +L IF+YY KPLGIKLPFYFDIVHFDQAA
Sbjct: 125 YEGLVELNLSCPNVPGKQIAYDFETDQLENIIFTYYTKPLGIKLPFYFDIVHFDQAAA 184

Query: 181 IFNKYPLAFINCNSIGNGLVIDDET VVIKPKNGFGGIGGDFIKPTALANVHAFYKRLNP 240
IFNKYPL+F+NCVNSIGNGLVI DE V+IKPKNGFGGIGGD+IKPTALANVHAFYKRL P
Sbjct: 185 IFNKYPLSFVNCVNSIGNGLVIKDEQVLKPKNGFGGIGGDYIKPTALANVHAFYKRLKP 244

Query: 241 SIKIIGTGGVKNGRDAFEHILCGASMVQIGTALQKEGPEIFQVRSRELKEIMADKGYQSL 300
SI IIGTGGVK GRDAFEHILCGASMVQIGTAL +EGP IF+RV++ELK IM +KGYQSL
Sbjct: 245 SIHIIGTGGVKTGRDAFEHILCGASMVQIGTALHQEGPAIFERVTKELKTIMVEKGYQSL 304

Query: 301 EDFRGQLNY 309
+DFRG L Y
Sbjct: 305 DDFRGNLRY 313

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1756

A DNA sequence (GBSx1863) was identified in *S.agalactiae* <SEQ ID 5457> which encodes the amino acid sequence <SEQ ID 5458>. This protein is predicted to be beta-lactam resistance factor. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4437(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB89121 GB:AJ277485 beta-lactam resistance factor
[Streptococcus pneumoniae]
Identities = 238/410 (58%), Positives = 304/410 (74%)

Query: 1 MALKELTAKEFESYSGNYDLQSFMTPEMAKLLKRGYDITYMGYQIDGKMEIISIVYTI 60

-1977-

MAL LT +EF++YS +SFMQ+ +M LL+KRG I Y+ + +G++++ ++VY++
 Sbjct: 1 MALTTLTKEEFQTYSDQVSSRSFMQSVQMGDLLEKRGARIVYLALKQEGEIQVAALVYSL 60
 Query: 61 PMTGGHMEVNSGPAHSNSKYLKHFYKELQNYAKSQGALELLIKPYDITYQEFTGEGKPKG 120
 5 PM GGLHME+NSGP ++ L FY EL+ YAK G LELL+KPY+TYQ F +G P
 Sbjct: 61 PMLGGLHMEVNSGPIYTQQDALPVFYAELKEYAKQNGVLELLVKPYETYQTDFDSQGNPID 120
 Query: 121 APNTYLIDDLTISIGYHHDGLHIGYPGGEPDWHYVKNLEGITPQNLLKSFSKKGRLVKKA 180
 A +I DLT +GY DGL GYPGGEPDW Y K+L +T ++LLKSFSKKG+PLVKKA
 10 Sbjct: 121 AEKKSIIQDLTDLGYQFDGLTTGYPGGEPDWLYYKDLTELSLLKSFSKKGKPLVKKA 180
 Query: 181 MSFGIKIRVLKREELHIFKDTISSTSDRRDYMDKSLDYYQDFYDSFGDKAEFVIATLNFR 240
 +FGI+++ LKREEL IFK+IT TS+RR+Y DKSL+YY+ FYD+FG++AEF+IA+LNF
 Sbjct: 181 ETFGIRLKKLKREELSIFKNITKETSERREYSKSLLEYEHFYDTFGEQAEFLIASLNF 240
 15 Query: 241 EYDHNQLNNAKLEEQITVLDNRHQNTDSAKYHRQTELNVQLASLDKRRKEVEPFQK 300
 +Y LQ KLEE + L N S K Q E +Q + + R+ E I+K
 Sbjct: 241 DYMSKLQGEQSKLEENLDKLRDLNPHSEKKQNQLREYSSQFETFEVRKAEARDLIEK 300
 20 Query: 301 FGNQDVVLGAGSLFIYSPKETVYLFSGSYTEFNKFIYAPAVLQEVVMQEAALKRQSTFYNFLG 360
 +G +D+VLGAGSLF+Y P+ET YLFSGSYTEFNKFIYAPA+LQ+YVM E++KR YNFLG
 Sbjct: 301 YGEEDIVLAGSLFVYMPQETTYLFSGSYTEFNKFIYAPALLQKYMLESIKRGIPKYNFLG 360
 Query: 361 IQGNFDGSDGVLRFKQNFNGYIVRKMGTFRYYPNPLKYKSIQLLKKILRR 410
 25 IQG FDGSDGVLRFKQNFNGYIVRK GTFRY+P+PLKYK+IQLLKKI+ R
 Sbjct: 361 IQGIFDGS DGVLRFKQNFNGYIVRKAGTFRYHPSPLKYKAIQLLKKIVGR 410

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5459> which encodes the amino acid sequence <SEQ ID 5460>. Analysis of this protein sequence reveals the following:

30 Possible site: 34
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2652(Affirmative) < succ>
 35 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 216/410 (52%), Positives = 291/410 (70%)
 40 Query: 1 MALKELTAKEFESYSGNYDLQSFMTPEMAKLLKKRGYDITYMGYQIDGKMEIISIVYTI 60
 MAL E++ ++F+ Y + SF+QT EMA L+ KRG ++G + DG++++ ++V++
 Sbjct: 1 MALIEISQEQFDHYCHSLVHHSFIQTSEMASLMAKRGAKPQFLGLEKDGELKVAAMVFSQ 60
 45 Query: 61 PMTGGHMEVNSGPAHSNSKYLKHFYKELQNYAKSQGALELLIKPYDITYQEFTGEGKPKG 120
 + GG ME+N+GP ++ + L+HFY +L++YAK + +EL++KPYD YQ F +G P
 Sbjct: 61 KVAGGWRMELNAGPNTNHPEELEHFFYTQLKDYAKQKDVIELILKPYDNYQSFDTDGIPIS 120
 50 Query: 121 APNTYLIDDLTISIGYHHDGLHIGYPGGEPDWHYVKNLEGITPQNLLKSFSKKGRLVKKA 180
 PNT LI LT++GY HDGL GYP GEP WHYVK LEGI L +SFSKKG+ L+KKA
 Sbjct: 121 RPNTDLISLLTALGYKHDGLKTGYPEGEPVWHYVKKLEGIDSSRLTRSFSGKKGKALIKKA 180
 Query: 181 MSFGIKIRVLKREELHIFKDTISSTSDRRDYMDKSLDYYQDFYDSFGDKAEFVIATLNFR 240
 +FGIK+R LKR+ELH FK+IT +TSDRRDY+DKSL YYQDFYDSFGD EF++ATLNF
 55 Sbjct: 181 NTFGIKLRQLKRDELHFFKEITEATSDRRDYLDKSLSYQDFYDSFGDSCEFMVATLNFE 240
 Query: 241 EYDHNQLNNAKLEEQITVLDNRHQNTDSAKYHRQTELNVQLASLDKRRKEVEPFQK 300
 +Y +NL+ +L I + N S K + EL +Q + R E F+++
 Sbjct: 241 DYLNLLKQRLQLATSINKVKGDLGKNPHSEKKQNRLKELSSQFETFQVRISEALHFLEE 300
 60 Query: 301 FGNQDVVLGAGSLFIYSPKETVYLFSGSYTEFNKFIYAPAVLQEVVMQEAALKRQSTFYNFLG 360
 +G +DV LAGSLFIY+ +E VYLFSGSY +FNKFIY+PA+LQE+ M +A+ + YNFLG
 Sbjct: 301 YGTDKDVFLAGSLFIYTEQEAIVYLFSGSYPKFNKFIYSPALLQEHAMLAHKGIKQYNFLG 360
 65 Query: 361 IQGNFDGSDGVLRFKQNFNGYIVRKMGTFRYYPNPLKYKSIQLLKKILRR 410

-1978-

I G FDGSDGVLRFKQNFNG+I++K GTFR YP P+KY I+L KK+L R
 Sbjct: 361 ITGKFDGSDGVLRFKQNFNGFILQKPGTFRCPYFPPIKYHFIRLAKKLLNR 410

A related GBS gene <SEQ ID 8895> and protein <SEQ ID 8896> were also identified. Analysis of this
 5 protein sequence reveals the following:

Homology to resistance proteins

The protein has homology with the following sequences in the databases:

57.4/74.9% over 409aa

Streptococcus

pneumoniae

GP|7649683| beta-lactam resistance factor Insert characterized

ORF01118(301 - 1530 of 1833)

GP|7649683|emb|CAB89121.1|AJ277485(1 - 410 of 410) beta-lactam resistance factor
 {Streptococcus pneumoniae}

%Match = 39.0

%Identity = 57.3 %Similarity = 74.9

Matches = 235 Mismatches = 103 Conservative Sub.s = 72

240 270 300 330 360 390 420 450
 IPVNRLLYKASNYVYALRKKRNS*LGKDTFMALKELTAKEFESYSGNYDLQSFMTPEMAKLLKKRGYDITYMGYQIDGK
 ||| || :||:| :|||:| ||:| | : : :
 MALTTLTKEEFQTYSDQVSSRSFMQSVQMGDLLEKRGARIVYLALKQEGE
 10 20 30 40 50

480 510 540 570 600 630 660 690
 MEIISIVYITPMTGGLHMEVNSGPAHSNSKYLKHFYKELQNYAKSQGALELLIKPYDTYQFTGEGKPKGAPNTYLIDDL
 :: :||:| |||||:| || : | || :| | ||||:| || :| | :| | :| |
 IQVAALVYSLPMLGGLHMEVNSGPIYTTQDALPVFYAELKEYAKQNGVLELLVKPYETVQTFDSQGNPIDAEKKSIIQDL
 60 70 80 90 100 110 120 130

720 750 780 810 840 870 900 930
 TSIGYHHDGLHIGYPGGECDWHYVKNLEGITPQNLKSFSSKGRPLVKKAMSFGIKIRVLKREELHIFKDITSSTSDRRD
 | :||: ||| ||||| || :| :| :||| ||||:| |||| :||:| |||| | ||:| | ||:| :
 TDLGYQFDGLTTGYPGGECDWLYYKDLTELTEKSLKSFSSKGRPLVKKAEFTFGIRLKKLREELSIFKNITKETSERRE
 140 150 160 170 180 190 200 210

960 990 1020 1050 1080 1110 1140 1170
 YMDKSLDYYQDFYDSFGDKAEFVIATLNFREYDHNQLNAKLEEQITVLDNRHQNTDSAKYHRQRTLVNQLASLDRK
 | ||||:| :||:| ||||:| || :| || |||| :| | | :| | :| : :|
 YSDKSLEYEYHFYDTFGEQAEFLIASLNFSDYMSKLQGEQSKLEENLDKLRDLKSNPHSEKKQNLREYSSQFETFEVR
 220 230 240 250 260 270 280 290

1200 1230 1260 1290 1320 1350 1380 1410
 RKEVEPFIQKFGNQDVVLGSLFIYSPKETVYLFSGSYTEFNKFYAPAVLQYVMQEALKRQSTFYXFLGIQGNFDGS G
 : | :||:| :||:| ||||:| ||:| ||||| ||||| ||||:| ||:| | ||||| ||||:|
 KAEARDLIEKYGEEDIVLAGSLFVYMPQETTYLFSGSYTEFNKFYAPALLQYVMLESIKRGIPKYNFLGIQGIQGNFDGSDG
 300 310 320 330 340 350 360 370

1440 1470 1500 1530 1560 1590 1620 1650
 VLXFKQNFNGYIVRKMGTFRYYPNPLKYKSIQLLKKILRRT*KISLHKLIFYAL*KASFISLLELLFIQTIMFVI*RNFIT
 || ||||| ||||:| ||||:| ||||:| ||||:|
 VLRFKQNFNGYIVRKAGTFRYHPSPLKYKAIQLLKKIVGR
 380 390 400 410

SEQ ID 8896 (GBS198) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
 extract is shown in Figure 26 (lane 6; MW 48.8kDa). It was also expressed in *E.coli* as a GST-fusion
 product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 6; MW 73.8kDa).

GBS198-GST was purified as shown in Figure 223, lane 4.

-1979-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1757

A DNA sequence (GBSx1864) was identified in *S.agalactiae* <SEQ ID 5461> which encodes the amino acid sequence <SEQ ID 5462>. This protein is predicted to be MurM protein. Analysis of this protein sequence reveals the following:

Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4418(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB89539 GB:AJ250767 MurM protein [Streptococcus pneumoniae]
Identities = 204/410 (49%), Positives = 286/410 (69%), Gaps = 17/410 (4%)

Query: 1 MYRE---ITAVEHDFVSESNQTNLLQSSNWPVKVDNWGSQLLGFFDGETQIASASILIK 57
MYR I +E+D+FV E N+LQSS W KVK +W + LG ++GE +A AS+LIK
Sbjct: 1 MYRYQIGIPTLEYDQFVKEHELANVLQSSAWKVKSDWNHERLGVYEGENLLAVASVLIK 60

Query: 58 SLPLGFSMLYIPRGPIMDYSNLDIVTKVLKDLKAFGKKQALFIKCDPLIYLK--MVNAK 115
SLPLG+ M YIPRGPI+DY + +++ VL+ +K++ + +RA+F+ DP I L +VN
Sbjct: 61 SLPLGYKMFYIPRGPILDYMDKELLKFVLQSIKSYARSKRAVFVTFDPSICLSQHLVN-- 118

Query: 116 DFENSPDEKEGLIAIDHLQAGADWIGRTTDLAHTIQPRFQANLYANQFGLDKMSKKTRQ 175
++ + E L ++ L + G W+G+IT++ TIQPR QA +Y F DK+SK TRQ
Sbjct: 119 --QDKREYPENLAIIVEILGQLGVKWSGQTTTMDDTIQPRIQAKIYKENFEEDKLSKSTRQ 176

Query: 176 AIRTSKNKGVDIQFGSHELLEDFAELMKKTEDRKGINLRGIDYYQKLLDTPNNSYITMA 235
AIRT++NKG++IQ+G ELL+ F+ELMKKTE RK I+LR YY+KLLD + +SYIT+
Sbjct: 177 AIRTARNKGLEIQYGGLELLDSFSELMKKTEKRKEIHLRNEAYVRKLLDNFKEDSYITLT 236

Query: 236 SLDVAKRLEKIEKECQIAQSERIKS--LELNREKKVKQHOGTIDRLNKEIDFLKEAQKAY 293
+LDV+KRL ++E+ Q+A+++ ++ E R KV+ + +RL +EIDFL +
Sbjct: 237 NLDVSKRLRELEE--QLAKNKALBAAFTSTRTSKVEAQKKEKERLVEEIDFL-QGYMMN 293

Query: 294 DRDIIPLAATLTLEFGNTSENIYAGMDDYFKSYSAPIYTWFFETAQRAFERGNIWQNMGGI 353
++ IPLAATL+LEFG TS N+YAGMDD FK Y+API TW+ETA+ AFERG +WQN+GG+
Sbjct: 294 EKSNIPLAATLSLEFGTTSVNLVYAGMDDDFKRYNAPILTWEYETARYAFERGMVWQNLGGV 353

Query: 354 ENDLGGGLYHFKSKFEPITIEEFIGEFNIPVN---RLLYKASNYVYALRKK 400
EN L+GGLYHFK KF P IEE++GEF +P + LL A ++ LRKK
Sbjct: 354 ENSLNGGLYHFKKFNPTIEEYLGEFTMPHTPLYPILLRLALDFRKTLRKK 403

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5463> which encodes the amino acid sequence <SEQ ID 5464>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2239(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 203/399 (50%), Positives = 274/399 (67%), Gaps = 4/399 (1%)

-1980-

Query: 5 ITAVEHDRFVSESNQITNLLQSSNWPVKVDNWGSQLLGFFDGETQIASASILIKSLPLGFS 64
 I+ EHD+FV Q LLQSS W KVKDNW + + F++ Q+A+A+ LI+ LPLGF+
 Sbjct: 13 ISPEEHDQFVLAQFQAGLLQSSKWGKVKDNWKKHERISFYENGQVAAAACLIRKLPFGT 72

Query: 65 MLYIPRGPIMDYSNLDIVTKVLKDLKAFGKKQRALFIKCDPLIYLMVNAKDFENSPDEK 124
 M+YIPRGPIMDY+N +++ V+K LK FGK +RALFIK DP + +K + + S +
 Sbjct: 73 MIYIPRGPIMDYANFELLDVFIKTLKTFGKSKRALFIKIDPSLVIKQT--LEGKESKEND 130

Query: 125 EGLIAIDHLQAGADWTGRTTDLAHTIQPRFQANLYANQFGLDKMSKKTRQAIRTSKNKG 184
 L I L++ G +W+GRT +L TIQPR QAN+YA F D + KK +Q+IRT+ NKG
 Sbjct: 131 VTLSIIAFLKKLGVESGRTKELEDTIQPRIQANIYAKDFDFDSLPPKAKQSIRTATNKG 190

Query: 185 VDIQFGSHELLEDFAELMKKTEDRKGINLRGIDYYQKLLDTPNNSYITMASLDVAKRLE 244
 V++ G ELL+DF+ LMKKTE+RKG LRG YYQKLL Y SYITMASLD+ ++ +
 Sbjct: 191 VNVITIGSELDDFSALMKKTENRKGII LRGSYYQKLLGIYAGQSYITMASLDLPEQKK 250

Query: 245 KIEKECQIAQSERIKSLELNREKKVKQHOGTIDRLNKEIDFLKEAQKAYDRDIPLAATL 304
 + ++ A +E+ + + ++ KV ++Q TI RL K++ L E Q A + IPLAATL
 Sbjct: 251 LLIQQLDKALAEQARLTDKSKPSKVAENQKTARLQKDLTILSE-QLATGQTRIPLAATL 309

Query: 305 TLEFGNTSENIYAGMDDYFKSYAPIYTWFTETAQRAFERGNIWQNMGGIENDLSGGLYHF 364
 TL +G TSEN+YAGMDD +++Y AP+ TW+ETA+ AF+RG W N+GG+EN GGLYHF
 Sbjct: 310 TLIYGETSENIYAGMDDYRNYQAPLLTWYETAKEAFKRGCRWHNLGGVENQQDGGLYHF 369

Query: 365 KSKFEPIIEEFIGEFNIPVNRLLYKASNYVYALRKKRNS 403
 K++ P IEEF GEFNIPV L+ + Y LRKK S
 Sbjct: 370 KARLNPTIEEFAGEFNIPVG-LVSSLAILTYNLRRKKLRS 407

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1758

A DNA sequence (GBSx1865) was identified in *S.agalactiae* <SEQ ID 5465> which encodes the amino acid sequence <SEQ ID 5466>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2669(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1759

A DNA sequence (GBSx1866) was identified in *S.agalactiae* <SEQ ID 5467> which encodes the amino acid sequence <SEQ ID 5468>. This protein is predicted to be beta-lactam resistance factor. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.07 Transmembrane 56 - 72 (55 - 74)

----- Final Results -----

-1981-

bacterial membrane --- Certainty=0.1829(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9625> which encodes amino acid sequence <SEQ ID 9626> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB89120 GB:AJ277484 beta-lactam resistance factor
 [Streptococcus pneumoniae]
 10 Identities = 166/410 (40%), Positives = 250/410 (60%), Gaps = 10/410 (2%)

Query: 6 MYHVTVGISEKEYDAFAIASSQTNLLHSSKWAQVKSNNWQNERLGFYKDDQLVAVASILIK 65
 MY +GI EYD F N+L SS W +VKSNNWQ+E+ G Y++++L+A ASILI+
 15 Sbjct: 1 MYRYQIGIPTLEYDQFVKEHELANVLQSSAWEEVKSNNWQHEKFGVYREEKLLATASILIR 60

Query: 66 SLPLGFTMLYIPRGPIMDYSNKELVNFVLKTLKNFGRKKRAVFAKFDPALLLRQYHLKEE 125
 +LPLG+ M YIPRGP+DY +KEL+NF +++++K++ R KRAVF FDP++ L Q + +E
 15 Sbjct: 61 TLPLGYKMFYIPRGPILDYGDKELLNFALQSIKSYARSKRAVFVTFDPSICLSQSLINQE 120

Query: 126 NVAEEIDESRQAIDNLKSAGAOWIGPTKAISETIQPRFOANIYTKANIEENFPKHTKRLI 185
 E E+ ID+L+ G +W G T+ + +TIQPR QA IY + E+ K TK+ I
 20 Sbjct: 121 KT--EFPENLAIDSLQQMGVRWSGKTEEMGDTIQPRIQAKIYKENFEEDKLSKSTKQAI 178

Query: 186 KDAKHRGVQIYRANIDDLPKFATVVALTENRKGVALRNENYFHQMLTIYGEDAYLYLAKV 245
 + A+++G++I ++ L F+ ++ TE RK + LRNE Y+ +L+ + + AY+ LA +
 25 Sbjct: 179 RTARNKGLEIQYGGLELLDSFSELMKKTEKRKEIHLRNEAYYKLLDNFKDKAYITLATL 238

Query: 246 NLPKRLAQFKEQLLQIQKDLSETPSHQKSRLTRLNQQEASVKQYILEFQEFSSKKYPD--- 302
 ++ KR + +EQL + + L ET + + +R +++ Q+ K+ +LE F ++Y D
 30 Sbjct: 239 DVSKRSQEELEQLAK-NRALEETFT-ESTRTSKVEAQKKE-KERLLEELTFLQYIDVQG 295

Query: 303 -EPVIAGILSIRFGNVLEMLYAGMDDSFRRKFYPQYLLNARVFEDAFKNDIVSANLGGVEG 361
 +A LS+ FG +YAGMDD F+++ L AF+ ++ NLGGVE
 35 Sbjct: 296 ARVPLAATLSLEFGTTSVNIYAGMDDDFKRYNAPILTWTYETARYAFERGMWQNLGGVEN 355

Query: 362 SLNDGLTKFKSNFNPMEFYIGEFNLAINPLLYKLANLAYTIRKKQRHSH 411
 SLN GL FK FNP EEY+GEF + +P LY L LA RK R H
 Sbjct: 356 SLNGGLYHFKEKFNPTIEEYLGFTMPHTP-LYPLRLALDFRKTLRKKH 404

- 40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5469> which encodes the amino acid sequence <SEQ ID 5470>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence
 45 INTEGRAL Likelihood = -0.32 Transmembrane 59 - 75 (59 - 75)

----- Final Results -----
 bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB89120 GB:AJ277484 beta-lactam resistance factor
 [Streptococcus pneumoniae]
 55 Identities = 166/402 (41%), Positives = 255/402 (63%), Gaps = 5/402 (1%)

Query: 9 KIGISEEEHDSFVKEHQQISVLQGSWDWAKIKNQWQNERIGIYKEEKQVASLSLLIKLLPL 68
 +IGI E+D FVKEH+ +VLQ S W ++K+ WQ+E+ G+Y+EEK +A+ S+LI+ LPL
 60 Sbjct: 5 QIGIPTLEYDQFVKEHELANVLQSSAWEEVKSNNWQHEKFGVYREEKLLATASILIRTLPL 64

Query: 69 GRSIIYIPRGPVMDYLDRLDVAFTMKTLKDYGKTKKALFIKYDPAILLKQYALGQEEEEK 128
 G + YIPRGP++DY D++L+ F +++++K Y ++K+A+F+ +DP+I L Q + QE+ E

-1982-

Sbjct: 65 GYKMFYIPRGPILDYGDKELLNFAIQSIKSYARSKRAVFTVDFPSICLSQS LINQEKTEF 124

Query: 129 PLALAAIKNLQEAGVHWTGLTMEIADSIQPRFQANIYTOENLEMQFPKHTRRLIKDAKQR 188
P LA I +LQ+ GV W+G T E+ D+IQPR QA IY + E + K T++ I+ A+ +

5 Sbjct: 125 PENLAIIDSLQQMGVRWSGKTEEMGDTIQPRIQAKIYKENFEEDKLSKSTKQAIRTARNK 184

Query: 189 GVKTYRVSQSELHKFSKIVSLTEKRKNISLRNEAYFQKLMTTYGDKAYLHLAKVNIPQKL 248
G++ L FS+++ TEKRR I LRNEAY++KL+ + DKAY+ LA +++ ++

10 Sbjct: 185 GLEIQYGGLELDSFSELMKKTEKRKEIHLRNEAYYKLLDNFKDKAXITLATLDVSKRS 244

Query: 249 DQYRQQLILINQDITRTQAHQKRLKKLEDQKASLERYITE---FEGFTDQYPEEVVAG 305
+ +QL N+ + T + R K+E QK ER + E + + D V +A

Sbjct: 245 QELEEQLAK-NRALEETFT-ESTRTSKVEAQKKEKERLLEELTFLQEYIDVGQARVPLAA 302

15 Query: 306 ILSISYGNVMEMLYAGMNDFFKKFYQYLLYPNVFQDAYQDGIWANMGGVEGSLDDGLT 365
LS+ +G +YAGM+DDFK++ L + + A++ G+IW N+GGVE SL+ GL

Sbjct: 303 TLSLEFGTTSVNIYAGMDDFKRYNAPILTWYETARYAFERGMWQNLGGVENSINGGLY 362

Query: 366 KFKANFAPTIEEFIGEFNLPVSPLYHIANTMYKIRKQLKNKH 407
FK F PTIEE++GEF +P PLY + RK L+ KH

20 Sbjct: 363 HFKEKFNPTIEEYLGEFTMPHTPLYPPLRLALDFRKTLRKKH 404

An alignment of the GAS and GBS proteins is shown below.

Identities = 226/407 (55%), Positives = 318/407 (77%), Gaps = 3/407 (0%)

25 Query: 5 LMYHVTVGISEKEYDAFAIASSQTNNLHSSKWAQVKSNNWQNERLGFYKDDQLVAVASILI 64
L ++ +GISE+E+D+F Q ++L S WA++K+ WQNER+G YK+++ VA S+LI

Sbjct: 4 LTFYAKIGISEEEHDSFVKEHQQISVLQSGSDWAKIKNQWQNERIGIYKEEKQVASLSLLI 63

30 Query: 65 KSLPLGFTMLYIPRGPIMDYSNKLVLNFKTLKNFGRKKRAVFAKDFPALLLRQYHLKE 124
K LPLG +++YIPRGP+MDY +++LV F +KTLK++G+ K+A+F K+DPA+LL+QY L +

Sbjct: 64 KLLPLGRSIIYIPRGPVMDYLDRLDVAFTMKTLDYDGKTKKALFIKYDPAILLKQYALGQ 123

Query: 125 ENVAAEIDESRQAIDNLKSAGAOWIGPTKAISSETIQPRFQANIYTKANIEENFPKHTKRL 184
E EE + AI NL+ AG W G T I+++IQPRFQANIYT+ N+E FPKHT+RL

35 Sbjct: 124 EE--EEKPLALAAIKNLQEAGVHWTGLTMEIADSIQPRFQANIYTOENLEMQFPKHTRRL 181

Query: 185 IKDAKHGQVQIYRANIDDLKPKFATVVALTENRKGVALRNENYPHQLMTIYGEDAYLYLAK 244
IKDAK RGV+ YR + +L KF+ +V+L+TE RK ++LRNE YF +LMT YG+ AYL+LAK

40 Sbjct: 182 IKDAKQRGVKTYRVSQSELHKFSKIVSLTEKRKNISLRNEAYFQKLMTTYGDKAYLHLAK 241

Query: 245 VNLPKRLAQFKEQLLQIQKDLSETPSHQSRRLTRLNQQEASVKQYILEFQEFSSKYPDEP 304
VN+P++L Q+++QL+ I +D++ T +HQK RL +L Q+AS+++YI EF+ F+ +YP+E

45 Sbjct: 242 VNIPQKLDQYRQQLILINQDITRTQAHQKRLKKLEDQKASLERYITEFEGFTDQYPEEV 301

Query: 305 VIAGILSIRFGNVLEMLYAGMDDSFRRKFYPQYLLNARVFEDAFKNDIVSANLGGVEGSLN 364
V+AGILSI +GNV+EMLYAGM+D F+KFYPQYLL VF+DA+++ I+ AN+GGVEGSL+

Sbjct: 302 VVAGILSISYGNVMEMLYAGMNDFFKKFYQYLLYPNVFQDAYQDGIWANMGGVEGSLD 361

50 Query: 365 DGLTKFKSNFNPMEFEEYIGEFNLAINPLLYKLANLAYTIRKKQRHSH 411
DGLTKFK+NF P EE+IGEFNL ++P LY +AN Y IRK+ ++ H

Sbjct: 362 DGLTKFKANFAPTIEEFIGEFNLPVSP-LYHIANTMYKIRKQLKNKH 407

SEQ ID 5468 (GBS377) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
55 extract is shown in Figure 65 (lane 4; MW 49kDa). It was also expressed in *E.coli* as a GST-fusion
product. SDS-PAGE analysis of total cell extract is shown in Figure 71 (lane 4; MW 74kDa).

GBS377-GST was purified as shown in Figure 212, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
vaccines or diagnostics.

-1983-

Example 1760

A DNA sequence (GBSx1867) was identified in *S.agalactiae* <SEQ ID 5471> which encodes the amino acid sequence <SEQ ID 5472>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 22
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
           bacterial cytoplasm --- Certainty=0.2073 (Affirmative) < succ>
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
10          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9627> which encodes amino acid sequence <SEQ ID 9628> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AAC76720 GB:AE000446 orf, hypothetical protein [Escherichia coli K12]
   Identities = 127/269 (47%), Positives = 189/269 (70%), Gaps = 1/269 (0%)

   Query: 7   SIKLVAVDIDGTLNLSKREITPEVAKAVQEAQSGVKIVATGRPIIGVQDLLEELKLINE 66
             +IKL+A+D+DGTL L I+P V A+ A+++GV +V+ TGRP GV + L+EL + +
20  Sbjct: 2   AIKLI AIDMDGTL L LDPHTISPAVKNAIAAARAGVNVVLT TGRPYAGVHNYLKEHMEQ 61

   Query: 67  EGDYVITFNGGLVQDTATGDDIIKETLT YEDYLD FELLARKLGVHMHAI TKEGIYTANRD 126
             GDY IT+NG LVQ A G + + L+Y+DY E L+R++G H HA+ + +YTANRD
25  Sbjct: 62  PGDYCITYNGALVQKAADGSTVAQTALS YDDYRFLEKLSREVGS HFHALDR TTTLYTANRD 121

   Query: 127 IGKTYTIEHVTLVNMPLFYRTPEEMG-DKEI IKLMMIDQPDILDAAIAKIPKKVLDNYTIV 185
             I YT+HE + +PL + E+M + + +K+MMID+P ILD AIA+IP++V + YT++
30  Sbjct: 122 ISYTYTHESFVATIPLVFCEAEKMDPNTQFLKVM MIDEPAILDQA IARIPQEVKEKYTVL 181

   Query: 186 KSTPFYLEILPKNVNKG TALLHLAEKMG LTV DQ TMAIGDEENDRAMLEVGNPVMQNGN 245
             KS P++LEIL K VNKG T + LA+ +G+ ++ MAIGD+END AM+E G V M N
35  Sbjct: 182 KSAPYFLEILDKRVNKG TGVKSLADVLGIKPEE IMAIGDQENDIAMIEYAGVGVAMDNAI 241

   Query: 246 PELKKIAKYITKSNEESGVAYALREWVIN 274
             P +K++A ++TKSN E GVA+A+ ++V+N
40  Sbjct: 242 PSVKEVANFVTKSNLEDGVAFAIEKYVLN 270

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3407> which encodes the amino acid sequence <SEQ ID 3408>. Analysis of this protein sequence reveals the following:

```

40  Possible site: 36
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
           bacterial cytoplasm --- Certainty=0.3474 (Affirmative) < succ>
45          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

50  Identities = 197/268 (73%), Positives = 235/268 (87%)

   Query: 7   SIKLVAVDIDGTLNLSKREITPEVAKAVQEAQSGVKIVATGRPIIGVQDLLEELKLINE 66
             SIKLVAVDIDGTL L R IT +V +AVQEAQ++GV +VIATGRPI GV LLE+L+LN
55  Sbjct: 2   SIKLVAVDIDGTL L TDDRRITDDV FQAVQEAQAGVHVVIATGRPIAGVISLLEQLNLN 61

   Query: 67  EGDYVITFNGGLVQDTATGDDIIKETLT YEDYLD FELLARKLGVHMHAI TKEGIYTANRD 126
             +G++VITFNGGLVQD TG++I+KE +TY+DYL+ E L+RKLGVHMHAI TKEGIYTANR+
60  Sbjct: 62  KGNHVITFNGGLVQDAETGEEIVKELMTYDDYLETEFLSRK LGVHMHAI TKEGIYTANRN 121

   Query: 127 IGKTYTIEHVTLVNMPLFYRTPEEMGDKEI IKLMMIDQPDILDAAIAKIPKKVLDNYTIVK 186

```

-1984-

IGKYT+HE TLVNMP+FYRTPEEM +KEIIK+MMID+PD+LDAAI +IP+ D YTIVK
 Sbjct: 122 IGKYTVHES TLVNMP IFYRTPEEMTKKEIIKMMIDEPDLLDAAIKQIPQHFDKYTIVK 181

Query: 187 STPFYLEILPKNVNKG TALLHLAEKMG LTVDTQ TMAIGDEENDRAMLEVVG NPFVVMQNGNP 246
 STPFYLE +PK V+KG A+ HLA+K+GL + Q TMAIGD ENDRAMLEV V NPVVM+NG P
 Sbjct: 182 STPFYLEFMPKTVSKGNAIKHLAKKLGLDMSQTMAIGDAENDRAMLEV VANPVMENGVP 241

Query: 247 ELKKIAKYITKSNEESGVAYALREWVIN 274
 ELKKIAKYITKSN +SGVA+A+R+WV+N
 Sbjct: 242 ELKKIAKYITKSNNDSGVVAHAIRKWVLN 269

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1761

- 15 A DNA sequence (GBSx1868) was identified in *S. agalactiae* <SEQ ID 5473> which encodes the amino acid sequence <SEQ ID 5474>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

- 20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2360(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 25 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07537 GB:AP001520 unknown conserved protein [Bacillus halodurans]
 Identities = 211/423 (49%), Positives = 285/423 (66%), Gaps = 5/423 (1%)

- 30 Query: 3 EKVFDPVHTYIYHVNNQVIYDLINTKEFQRLRRIKQTSTTSFTFHGAESRFSHCLGVYE 62
 EKVF+DPVH YIHV +++I+ LI TKEFQRLRR++Q TT TFHGAEH+RF+H LGVYE
 Sbjct: 12 EKVFDPVHRYIYHVRDELIALIGTKEFQRLRRVRQLGTTFLTFHGAETR FNHSLGVYE 71
- 35 Query: 63 LARKVTEIFDEHYS DLWNKNESLLTMAAALLHDIGHGAYSHTFERLFNTDHEAYTQEIIIT 122
 + R++ E+ F WN+ E LLT+ AALLHDIGHG +SH+FE++F+TDHE +T+ +I
 Sbjct: 72 ITRRIIEVFQGR--PYWNEEERLLTLCAALLHDIGHGPFSSHSEKVFDTDHEEWTRRMIV 129
- 40 Query: 123 NPTEINAILRKVAPDFDPKVASVINHSYPNKKVQVQLISSQIDCDRMDYLLRDSYYTAAS 182
 T EI+ +L K+ DFP KVA VI +YPNK V +ISSQID DRMDYL RD+YYT S
 Sbjct: 130 GDT-EIHNVLLKMGDDFPQKVADVIEKTYPNKLVTSIISSQIDADRMDYLQRDAYYTGVS 188
- 45 Query: 183 YGQFDLTRLRVRPTDSGIAFARNGMHAVEDYIVSRFQMYMQVYFHPASRAMELLQNL 242
 YG FD+ RILRV+RP + + ++GMHAVEDYI+SR+QMY QVYFHP +R+ E++L +
 Sbjct: 189 YGHFDMERILRVMRPMEDQVVIKQSGMHAVEDYIMSRYQMYQVYFHPVTRSAEVLISKV 248
- 50 Query: 243 LKRARFLFDTHRDFFEQTSPNLIPFFTDQYDLQDYLA LDDGVMNTYFQSWMQADDNILD 302
 KR + L++ F+Q + F L DYL LD+ + YFQ W + +D IL+D
 Sbjct: 249 FKRVKDLYEQGYK-FKQEPKH FYSLFEGNMSLDDYLR LDESITMYFQI WQEEDRILSD 307
- 55 Query: 303 LANRFINRKVKFSITFEESDKEN-LVKMKELVSQVGFDPDYTG VHANFDLPYDVYRPEH 361
 L RFINR++FK I F + + N ++++L +Q DP+YY V ++ DLPYD YRP
 Sbjct: 308 LCVRFINRQLFKYIEFNPNLQMNDWPRLQQLFAQAEIDPEYYLVVDSSSDLPYDFYRPG 367
- Query: 362 SNPRTEIQIIQKNGQLAELSSLSPIVKALTGSNYGDQRFYFPKEMLTLDLFSSTKEEFQ 421
 R I +I NG+L ELS S +V+A++G D + YFP + LT S K+E
 Sbjct: 368 EEERLPIHLIMPNGKLRELSRESVDVEAISGKKRTDHKLYFPMDC L TDQSDHKEIKQEIL 427
- Query: 422 SYI 424
 S +
 Sbjct: 428 SLL 430

60

-1985-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5475> which encodes the amino acid sequence <SEQ ID 5476>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2220(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 321/428 (75%), Positives = 379/428 (88%)

Query: 1 MNEKVFRDPVHTYIHVNNQVIYDLINTKEFQRLRRIKQTSTTSFTFHGAHSRFSHCLGV 60

MNEKVFRDPVH YIH++N +IYDLINTKEFQRLRRIKQ TT+FTFHGAHSRFSHCLGV

Sbjct: 1 MNEKVFRDPVHNYIHIDNPLIYDLINTKEFQRLRRIKQVPTTAFTFHGAHSRFSHCLGV 60

Query: 61 YELARKVTEIFDEHYSDLWNKESLLTMAAALLHDIGHGAYSHTFERLNTDHEAYTQEI 120

YE+AR+VT IF+E Y+D+WNK+ESL+TM AALLHDIGHGAYSHTFE LF+TDHEA+TQEI

Sbjct: 61 YEIARRVTAIFEEKYADIWNKDESLVTMTAALLHDIGHGAYSHTFEVLFTDHEAFTQEI 120

Query: 121 ITNPTEINAILRKVAPDFDPKVASVINHSYPNKQVVQLISSQIDCDRMDYLLRDSYTA 180

ITNP TEINAIL + APDFDPKVASVINH+YPNKQVVQLISSQIDCDRMDYLLRDSY++A

Sbjct: 121 ITNPETEINAILVRHAPDFDPKVASVINHTYPNKQVVQLISSQIDCDRMDYLLRDSYFSA 180

Query: 181 ASYGQFDLTRLRVRIRPTDSGIAFARNGMHAVEDYIVSRFQMYMQVYFHPASRAMELLQ 240

A+YGQFDL RILRVIRP + GI F +GMHAVEDYIVSRFQMYMQVYFHPASRA+EL+LQ

Sbjct: 181 ANYGQFDLMRILRVIRPVEDGIVFEHSGMHAVEDYIVSRFQMYMQVYFHPASRAVELILQ 240

Query: 241 NLLKRARFLFDTHRDFFEQTSPNLIPIFFTDQYDLQDYALDDGVMNTYFQSWMQADDNII 300

NLLKRA+ L+ + +F++T+P LIPFF + +L DY+ALDDGVMNTYFQ WM ++D+IL

Sbjct: 241 NLLKRAQHLYPEQQAYFQKTAPGLIPFFEKKANLADYIALDDGVMNTYFQVWMASEDHIL 300

Query: 301 ADLANRFINRKVFKSITFEESDKENLVKMKELVSQVGFDPDYTYGVHANFDLPYDVYRPE 360

+DLA+RFINRK+ KS+TF++ + L +++++LV VGFDPDYTYG+H NFDLPYD+YRPE

Sbjct: 301 SDLASRFINRKILKSVTFDQDSQGELERLRQLVESVGFDPDYTYGIHINFDLPYDIYRPE 360

Query: 361 HSNPRTEIQIIQKNGQLAELSSLSPIVKALTGSGNYGDRFYFPKEMLTLDLSLFSSTKEEF 420

NPRT+I+++QK+G LAELS LSPIVKALTG+ YGD+RFYFPKEML LD LF+ +KE F

Sbjct: 361 LENPRTQIEMMQKDGSLAELSLSPIVKALTGTTYGDRRFYFPKEMLELDDLAFAPSKETF 420

Query: 421 QSYITNEH 428

SYI+N H

Sbjct: 421 MSYISNGH 428

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1762

A DNA sequence (GBSx1869) was identified in *S.agalactiae* <SEQ ID 5477> which encodes the amino acid sequence <SEQ ID 5478>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4789(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

-1986-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5479> which encodes the amino acid sequence <SEQ ID 5480>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3650(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 64/127 (50%), Positives = 89/127 (69%)

Query: 5 MKLEINNNIQIDNETEMIHEIHDCQFIEKGSYVYLYNYINAEGERVVIKANHEELIMTRFS 64

MKL++ N+I+ +ETE+I EIHC++ EKG Y YL Y N + E+VVIK N EL M+RFS

Sbjct: 1 MKLQLTNHIRFGDETEIIQEIHCCEWREKGGYQYLIYQNTDKEKVVIKYNETELTMSRFS 60

Query: 65 NPKSVMRFHRETPALVNIPTPLGVQHLITETSHYQFDLSQQLRHINYVLKQTETGDCFAN 124

NP+S+M+F L+ +PTP+GVQ +T+TSHY D S Q+L ++Y L Q +T FA+

Sbjct: 61 NPQSIMKFFAGKKVLIALPTPMGVQQFLTDTSHYHLDSCQKLDLHYHLLQAQTEMLFAS 120

Query: 125 YELRIQW 131

Y L + W

Sbjct: 121 YHLELSW 127

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1763

A DNA sequence (GBSx1870) was identified in *S.agalactiae* <SEQ ID 5481> which encodes the amino acid sequence <SEQ ID 5482>. This protein is predicted to be cation-transporting ATPase PacL (ctpF). Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -13.27 Transmembrane 256 - 272 (246 - 276)

INTEGRAL Likelihood = -9.02 Transmembrane 64 - 80 (58 - 85)

INTEGRAL Likelihood = -8.49 Transmembrane 833 - 849 (828 - 855)

INTEGRAL Likelihood = -8.17 Transmembrane 89 - 105 (81 - 107)

INTEGRAL Likelihood = -7.48 Transmembrane 864 - 880 (860 - 884)

INTEGRAL Likelihood = -3.29 Transmembrane 287 - 303 (284 - 306)

INTEGRAL Likelihood = -2.55 Transmembrane 754 - 770 (753 - 773)

INTEGRAL Likelihood = -0.85 Transmembrane 695 - 711 (694 - 711)

INTEGRAL Likelihood = -0.75 Transmembrane 793 - 809 (792 - 809)

----- Final Results -----

bacterial membrane --- Certainty=0.6307(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13439 GB:Z99112 similar to calcium-transporting ATPase

[Bacillus subtilis]

Identities = 380/888 (42%), Positives = 545/888 (60%), Gaps = 49/888 (5%)

Query: 10 FYTQGGQEVLTSLESS-REGLSTTEAKNRLEMYGRNELEEGKKRSLIAKFFDQFKDLMII 68

F+ GQ ++L + +S ++GL+ E K RL+ +G NEL+EGKK S + FF QFKD M++

Sbjct: 3 FHEMGQTDLLEATNTSMKQGLTEKEVKKRLDKHGPNELOEGKKTSALLLFFAQFKDFMVL 62

Query: 69 ILLVAAALSVITEGMHG-LTDALIIILAVVILNAAFVYQEGQAEAAIEALKDMSSPIARV 127

-1987-

+LL A +S G G DA+ I+A+V +N G +QE +AE +++ALK++S+P
 Sbjct: 63 VLLAATLIS----GFLGEYVDAVAIIAIVFVNGILGFFQERRAEQSLQALKELSTPHVMA 118
 Query: 128 RRDGHTIEVDSKELVPGDLVLMLEAGDVPADLRILEAASLKIEEAALTGESVPVEKDISQ 187
 R+G ++ SKELVPGD+V +GD + AD+R++EA SL+IEE+ALTGES+PV K +
 Sbjct: 119 LREGSWTKIPSKELVPGDIVKFTSGDRIGADVRIVEARSLEIEESALTGESIPVVKHADK 178
 Query: 188 VVAEDAGIGDRVNMAQNSNVTYGRGYGVVNTGMYTEVGKIADMLANADESETPLKQSL 247
 + D +GD NMA+ + VT G G GVV TGM T +GKIADML +A TPL++ L
 Sbjct: 179 LKKPDVSLGDITNMAFMGTIVTRSGVGVVVGTMNTAMGKIADMLSESACTLSTPLQRRL 238
 Query: 248 VQLSKLLTYLIVIIAVITFLVGIFVRKEGWIEGLMTSVALAVAAIPEGLPAIVTIVLSMG 307
 QL K+L + +++ V+ VG+ ++ + V+LAVAAIPEGLPAIVT+ LS+G
 Sbjct: 239 EQLGKILIVALLLTVLVAVGV- IQGHDLYSMFLAGVSLAVAAIPEGLPAIVTVALSLG 297
 Query: 308 TKTLAKRNSIVRKLPVETLGTSTIIASDKTGTILTMNQMTVEKVT----- 353
 + + K+ SIVRKLPVETLG II SDKTGT+T N+MTV V++
 Sbjct: 298 VQRMKQKSIVRKLPVETLGCASIIICDKTGTMTQNKMTVTHVWSSGGKTWRVAGAGYEP 357
 Query: 354 NGVLQSSSEISVDNNTL-----RIMNFSNDTKIDPSGKLIGDPTETALVQFGLDKN 405
 G + +EISV+ + + N SN K D L.GDPTE AL+
 Sbjct: 358 KGSFTLNEKEISVNEHKPLQQMLLFGALCNSNIEKRDGEYVLDGDPTEGALLTAARKGG 417
 Query: 406 FDVREVLKNEPRVAELPFDSDRKLMSTIHKESDGRYFIAVKGAPDQLLKRVTKIEDNGLV 465
 F V N + E PFDS RK+M+ I + D + +I KGAPD L++R ++I +G
 Sbjct: 418 FSKEFVESNYRVIEEFPFDSARKMMTVIVENQDRKRYIITKGAPDVLQMQRSSRIYYDGSA 477
 Query: 466 RDITAEDEKAILNTNKEKALQALRVLMAYK--YETQIPSLETDIVESDLVFSGLVGMID 523
 + E K + LA QALR + +AY+ + PS+E E DL GL G+ID
 Sbjct: 478 ALFSNERKAETEAVLRHLASQALRTIAVAYRPIKAGETPSMEQ--AEKDLTMLGLSGIID 535
 Query: 524 PERPEAAEA VRVAKEAGIRPIMITGDHQDTAEAIKRLGIIDANDTEDHVFTGAELNELS 583
 P RPE +A++ +EAGI+ +MITGDH +TA+AIK L ++ + + G LNELS
 Sbjct: 536 PPRPEVRQAIKECREAGIKTVMITGDHVTAKAIKDLRLPKS--GKIMDGKMLNELS 592
 Query: 584 DEEFQKVFQYSVYARVSPPEHKVRIVKAWQNDGKVVMATGDGVNDAPSLKTADIGIGMGI 643
 EE V + V+ARVSPPEHK++IVKA+Q +G +VAMTGDGVNDAP++K ADIG+ MGI
 Sbjct: 593 QEELSHVVEDVYVFARVSPPEHKLKIVKAYQENGHIVAMTGDGVNDAPAIKQADIGVSMGI 652
 Query: 644 TGTEVSKGASDMVLADDNFATIIVAVEEGRKVFNSIQKSIQYLLSANMAEVFTIFFATLL 703
 TGT+V+K AS +VL DDNFATI A++EGR ++ NI+K I+YLL++N+ E+ + FA LL
 Sbjct: 653 TGTDVAKEASSLVLDNDFATIKAISKEGRNIYENIRKFIYLLASNVGEILVLMFAMLL 712
 Query: 704 GWDV-LAPVHLLWINLVTDTLPAIALGVPEPAEPGVMTHKPRGRQSNFFDGGVGMGAIYQG 762
 + L P+ +LW+NLVTD LPA+ALG++ E VM KPR + F + ++ +G
 Sbjct: 713 ALPLPLVPIQILWVNLVTDGLPAMALGMDQPEGDVMKRKPRHPKEGVFARKLGWKVVSRG 772
 Query: 763 ILQITILVLGVYGWALMY---PEHAGYRMIHADALTMAFATLGLIQLVHAFNVKSVYQSIF 819
 L I V + + ++Y PE+ Y A T+AFATL L QL+H F+ +S S+F
 Sbjct: 773 FL--IGVATILAFIIVYHRNPENLAY-----AQTIATLVLVLAQLIHVFDCRS-ETSVF 823
 Query: 820 TVGAFKNRTFNWSIPVAFILLMVTIVVPGFNKLFHVTHLSSTQWLTVV 867
 + F+N ++ + +L+V I P +FH ++ W+ V+
 Sbjct: 824 SRNPFQONLYLIGAVLSSILLMLVVIYYPPLQPIFHTVAITPGDWMLVI 871

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4171> which encodes the amino acid sequence <SEQ ID 4172>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

60	INTEGRAL	Likelihood = -12.47	Transmembrane	863 - 879 (856 - 883)
	INTEGRAL	Likelihood = -10.08	Transmembrane	64 - 80 (58 - 86)
	INTEGRAL	Likelihood = -8.97	Transmembrane	256 - 272 (249 - 275)
	INTEGRAL	Likelihood = -8.55	Transmembrane	89 - 105 (81 - 107)
	INTEGRAL	Likelihood = -5.84	Transmembrane	832 - 848 (827 - 850)
65	INTEGRAL	Likelihood = -3.13	Transmembrane	287 - 303 (284 - 307)
	INTEGRAL	Likelihood = -2.66	Transmembrane	762 - 778 (761 - 779)

-1988-

INTEGRAL Likelihood = -0.37 Transmembrane 685 - 701 (685 - 701)

----- Final Results -----

5 bacterial membrane --- Certainty=0.5989(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 735/892 (82%), Positives = 813/892 (90%), Gaps = 1/892 (0%)

10 Query: 3 KEQKSLFYTQGGEEVLTSLESSREGLSTTEAKNRLEMYGRNELEEGKKRSLIAKFFDQF 62
 KEQ+ FYTQ +E VL LE+SREGL++ +AK RL YGRNEL+EG+KRS L KF DQF
 Sbjct: 3 KEQRHEAFYTQSEETVLAQLETSREGLTSAQAKERLAEYGRNELDEGEKRS LFMKF L DQF 62

15 Query: 63 KDLMI ILLVAAALSVITEGMHGLTDAL I I L AV I L N A A F G V Y Q E G Q A E A A I E A L K D M S S 122
 KDLMI I L + V A A L S V + T E G M G L T D A + I I L A V I L N A A F G V Y Q E G Q A E A A I E A L K M S S
 Sbjct: 63 KDLMI I L I V A A L L S V L T E G M E G L T D A I I L A V I L N A A F G V Y Q E G Q A E A A I E A L K M S S 122

20 Query: 123 PIARVRRDGH TIEVDSKELVPGD L V M L E A G D V V P A D L R L L E A A S L K I E E A A L T G E S V P V E 182
 P+AR+RRDGH E+DSKELVPGD+V+LEAGD V V P A D L R L L E A S L K I E E A A L T G E S V P V E
 Sbjct: 123 PLARIRRDGHVTEIDSKELVPGD I V L L E A G D V V P A D L R L L E A N S L K I E E A A L T G E S V P V E 182

25 Query: 183 KDISQVVAEDAGIGDRVNMAYQNSNVTYGRGYGVVINTGMYTEVGKIADMLANADESETP 242
 KD+S V+EDAGIGDRVNM YQNSNVTYGRG GV+INTGMYTEVG IA MLANADE++TP
 Sbjct: 183 KDLSTAVSEDAIGIGDRVNMGYQNSNVTYGRGIGVITNTGMYTEVGHIAGMLANADETDTP 242

30 Query: 243 LKQSLVQLSKLLTYLIVIIAVITFLVGIFVRKEGWIEGLMTSVALAVAAIPEGLPAIVTI 302
 LKQ+L LSK+LTY I++IA +TF VG+F+R + +EGLMTSVALAVAAIPEGLPAIVT+
 Sbjct: 243 LKQNL D N L S K I L T Y A I L V I A A V T F A V G V F L R G Q H P L E G L M T S V A L A V A A I P E G L P A I V T V 302

35 Query: 303 VL S M G T K T L A K R N S I V R K L P A V E T L G S T E I I A S D K T G T L T M N Q M T V E K V Y T N G V L Q S S S E 362
 VLS+GT+ LAKRN+I+RKLP A V E T L G S T E I I A S D K T G T L T M N Q M T V E K V Y T N G L Q S S S
 Sbjct: 303 VLSLGTQVLAKRNA I R K L P A V E T L G S T E I I A S D K T G T L T M N Q M T V E K V Y T N G T L Q S S S A 362

40 Query: 363 EISVDNNTLRIMNFSNDTKIDPSGKLIGDPTETALVQFGLDKNFDVREVLKNEPRVAELP 422
 +I+ DN TLR+MNF+NDTK+DPSGKLIGDPTETALV+FGLD NFDVRE + EPRVAELP
 Sbjct: 363 DIAFDNTTLRVMNFANDTKVDPGKLIGDPTETALVEFGLDHNFDVREAMVAEPRVAELP 422

45 Query: 423 FDSDRKLMSTIHKESDGRYFIAVKGAPDQLLKRVTKIEDNGLVRDITAEDKEAILNTNKE 482
 FDSDRKLMSTIHK++DG+YFIAVKGAPDQLLKRV T+IE+NG +R IT DK+ IL+TNK
 Sbjct: 423 FDSDRKLMSTIHKQADGKYFIAVKGAPDQLLKRV TQIEENGQIRPITDADKKTILDTNKS 482

50 Query: 483 LAKQALRVLM MAYKYETQIP SLETDIVESDLVFSGLVGMIDPERPEAAEAVRVAKEAGIR 542
 LAKQALRVLM MAYKY +P+LET+IVE++LVFSGLVGMIDPERPEAA+AV+VAKEAGIR
 Sbjct: 483 LAKQALRVLM MAYKYS DALPTLETEIVEANLVFSGLVGMIDPERPEAAQAVKVAKEAGIR 542

55 Query: 543 PIMITGDHQDTAEAI AKRLGIIDANDTEDHVFTGAELNELSDEEFQKVFQYSVYARVSP 602
 PIMITGDHQDTA+AI AKRLGII+ D DHVFTGAELNELSDEEFQKVFQYSVYARVSP
 Sbjct: 543 PIMITGDHQDTAKAI AKRLGII E-EDGV DHVFTGAELNELSDEEFQKVFQYSVYARVSP 601

60 Query: 603 EHKVRIVKAWQNDGKVVAMTGDGVNDAPS LKTADIGIGMGITGTEVSKGASDMVLADDNF 662
 EHKVRIVKAWQ N+GKVVAMTGDGVNDAPS LKTADIGIGMGITGTEVSKGASDMVLADDNF
 Sbjct: 602 EHKVRIVKAWQNEGKVVAMTGDGVNDAPS LKTADIGIGMGITGTEVSKGASDMVLADDNF 661

65 Query: 663 ATIIIVAVEEGRKVF SNIQKSIQYLLSANMAEVFTIFFATLLGWDVLAPVHLLWINLVTD T 722
 ATIIIVAVEEGRKVF SNIQK+IQYLLSANMAEVFTIF ATL GWDVL PVHLLWINLVTD T
 Sbjct: 662 ATIIIVAVEEGRKVF SNIQK T I QYLLSANMAEVFTIFLATLFGWDVLQPVHLLWINLVTD T 721

Query: 723 LPAIALGV E P A E P G V M T H K P R G R Q S N F F D G G V M G A I I Y Q G I L Q T I L V L G V Y G W A L M Y P E H 782
 LPAIALGV E P A E P G V M H K P R G R + S + F F D G G V A I + Y Q G Q T I L V L G V Y G + A L M + P E H
 Sbjct: 722 LPAIALGV E P A E P G V M K H K P R G R K S S F F D G G V K E A I L Y Q G A F Q T I L V L G V Y G F A L M F P E H 781

Query: 783 AGYRMIHADALTMAFATLGLIQLVHAFNVKSVYQSIFTVGAFKNRTFNWSIPVAFILLMV 842
 Y +HADALTMA+ TLGLIQLVHA+NVKSVYQSIFTVG FKN+ FN+SIPVAF+ LM
 Sbjct: 782 TSYHDVHADALTMAVVTGLIQLVHAYNVKSVYQSIFTVG L F K N K L F N Y S I P V A F V A L M A 841

Query: 843 TIVVPGFNKLFHVTHLSSTQWLTVVIGSLLMVVLTEIVKFIQRKLGQDEKAI 894

-1989-

T+VVPGFN+ FHVTHL+ TQWL V+IGSLLMVVL E+VK +QR LGQDEKAI
 Subject: 842 TVVVPGFENOFFHVTHLTITOWLVVIIGSLLMVVLVELVKAVORS LGODEKAI 893

A related GBS gene <SEQ ID 8897> and protein <SEQ ID 8898> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6
McG: Discrim Score: -9.88
GvH: Signal Score (-7.5): -6.96
 Possible site: 14

>>> Seems to have no N-terminal signal sequence

ALOM program count: 9 value: -13.27 threshold: 0.0

INTEGRAL	Likelihood = -13.27	Transmembrane	256 - 272 (246 - 276)
INTEGRAL	Likelihood = -9.02	Transmembrane	64 - 80 (58 - 85)
INTEGRAL	Likelihood = -8.49	Transmembrane	833 - 849 (828 - 855)
INTEGRAL	Likelihood = -8.17	Transmembrane	89 - 105 (81 - 107)
INTEGRAL	Likelihood = -7.48	Transmembrane	864 - 880 (860 - 884)
INTEGRAL	Likelihood = -3.29	Transmembrane	287 - 303 (284 - 306)
INTEGRAL	Likelihood = -2.55	Transmembrane	754 - 770 (753 - 773)
INTEGRAL	Likelihood = -0.85	Transmembrane	695 - 711 (694 - 711)
INTEGRAL	Likelihood = -0.75	Transmembrane	793 - 809 (792 - 809)
PERIPHERAL	Likelihood = 1.06	714	

modified ALOM score: 3.15

*** Reasoning Step: 3

----- Final Results -----

```

bacterial membrane --- Certainty=0.6307(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

ORF01112 (328 - 2901 of 3282)

EGAD|108247|BS1566(3 - 871 of 890) hypothetical protein {Bacillus subtilis} OMNI|NT01BS1841
cation-transporting ATPase PacL GP|2337795|emb|CAA74269.1||Y13937 putative PacL protein
{Bacillus subtilis} GP|2633938|emb|CAB13439.1||Z99112 similar to calcium-transporting
ATPase {Bacillus subtilis} PIR|H69877|H69877 calcium-transporting ATPase homolog yloB -
Bacillus subtilis

```
%Match = 29.0
```

```
%Identity = 43.9    %Similarity = 64.5
```

Matches = 376 Mismatches = 291 Conservative Sub.s = 176

249 279 309 339 369 396 426 456
GVVINS~~ET~~CFHK~~NR~~SLFVCGETKGGKVL~~L~~KEOKKSLFYTOGOEEVLTSLESS-REG~~L~~STTEAKNRLEMYGRNELEEGKKR

| : || :: : :| ::||: | | ||: :| |||:||||
 MKFHMGQTDLLLEATNTSMKQGLTEKEVKRKRLDKHGPNELQEGKKT
 10 20 30 40

486 516 546 576 606 636 666 696
SLIAKFDFDPKDLMIITILLVAAALSVITEGMHGLTDLALIIILAVVIILNAAPGVYOEEOAEAAIEALKDMSSPIARVRRDGH

| : || |||:|::|| |::: | | :|:| :| :| :|| :|| ::|||::| | :|
 SALLLFFAQPKDFMVLVLL---AATLISGFLGEYDVAIIAIVFVNGILGFFQERRAEQSLQALKELSTPHVMALREGS
 60 70 80 90 100 110 120

726 756 786 816 846 876 906 936
TIEVDSKELVPGDLVMLEAGDVVPADIRLLLEAASLKIEEAAALTGESVPEVKDISOVVAEDAGIGDRVNMAYONSNTVTYGR

WTKIPSKELVPGDIVKFTSGDRIGADVRIEARSLEIEESALTGESIPVVKHADLKKPDVSLGDTITNMAFMGTIVTRGS

966 996 1026 1056 1086 1116 1146 1176

GYGVVITNGMYTEVGKTDMLANADESETPLKOSLVOLSKLLTYLLIVIIAVITFLVGI FVRKEGWIEGLMTSVALAVAAI

GVGVVVGTMNTAMGKIADMLESAGTLSTPLQRRLEQLGKILLVALLLVVLVAVGV-IQGHDLYSMFLAGVSLAVAAI
 220 230 240 250 260 270 280

[illegible]

[illegible]

Example 1764

```

Possible site: 48
55  >>> Seems to have no N-terminal signal sequence

```

60

The protein has homology with the following sequences in the GENPEPT database.

-1991-

>GP:CAB48940 GB:AJ248283 hypothetical protein [Pyrococcus abyssi]
Identities = 60/221 (27%), Positives = 100/221 (45%), Gaps = 37/221 (16%)

Query: 33 KIDHLHIA-----GDISNHFTKDTLP-FINNKKH--IKLSYNLGNHMDLDELTE--TE 80
KID L I GD+SN+ D + I+ L + L GNHD+ L +
Sbjct: 15 KIDVLKIPDIAIQGLDLSNYGEPDIIENLISELVTQLDPVPLLVIPGNHDIYGLNDIFAA 74

Query: 81 IQRLDFQTYR-----FDDKMLLAFHGWYDYSFSNN--RDIKDVEKLKKTFWFD 126
QR + R ++ ++ GWYDYS + KD ++K F F
Sbjct: 75 FQRFNKLVKRAGAIPMEGPLILEEIGIVGVPGWYDYS LAPGYLNMTKDEYEIK-AFGFR 133

Query: 127 RR-----LKRPNNDVTIQASILKRLDEILAKVDSS--NIIAMHFVPHKQFTMT--HPRF 177
R +K +D + L L++ ++++ S ++I+A+HF P K +P
Sbjct: 134 RLEDADYIKSSLSDEELVRWNLNLEKFISEIRESVNDVILALHFAPFKDSLKYTG NPEI 193

Query: 178 SPFNAFLGSQAYHDLFQKYHIKDVVFGHAHRSFGDVKIGET 218
F+A++GSQ + + +++I +V GH HRS + IG+T
Sbjct: 194 DYFSAYMGSRFGEFALRHNI GLIVHGHTHRSI-EYYIGKT 233

20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1765

A DNA sequence (GBSx1872) was identified in *S.agalactiae* <SEQ ID 5485> which encodes the amino acid sequence <SEQ ID 5486>. Analysis of this protein sequence reveals the following:

Possible site: 44
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -2.18 Transmembrane 173 - 189 (173 - 189)

----- Final Results -----
bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB16056 GB:Z99124 fructose-1,6-bisphosphatase [Bacillus subtilis]
Identities = 314/642 (48%), Positives = 446/642 (68%), Gaps = 7/642 (1%)

Query: 2 SNFYKLLKEKFPKEDIVTEMINLEAICQLPKGTEYFISDLHGEYDAVDYLLRTGAGSIR 61
S + LL +K+ +E +VTE+INL+AI LPKGTE+F+SDLHGEY A ++LR G+G ++
Sbjct: 33 SKYLDLLAQKYDCEEKVVTETIINLKAILNLPKGTEHFVSDLHGEYQAFQHVLNNGSGRVK 92

Query: 62 AKLLDCFDWQKIVAVDLDDFCILLYPKEKLAFDKMNLSASAYKTKLW-EMIPLQIQVLK 120
K+ D F I ++D+ L+YYP++KL K + A + + E I I+++
Sbjct: 93 EKIRDIFSGV-IYDREIDELAALVYYPEDKCLKLIKHDFAKEALNEWYKETIHRMIKLV 151

Query: 121 YFSSKYTKSKVRKQLSGKFAYIIEELLAEIDRNPEKSYFDTIIEKLFELDQVEDLIIVL 180
Y SSKYT+SK+RK L +FAYI EELL + ++ K+ Y+ II+++ EL Q + LI L
Sbjct: 152 YCSSKYTRSKLRKALPAQFAYITEELLYKTEQAGNKEQYSEIIDQIIELGQADKLITGL 211

Query: 181 SQTIVQLIIDHLHVVGDIYDRGRYPDRILNRLMAFPNLDIQWGNHVDVTWMGAASGSYLCM 240
+ ++Q L++DHLHVVGDIYDRG PDRI+ L+ + ++DIQWGNHDV W+GA SGS +C+
Sbjct: 212 AYSVQRLVVDHLHVVGDIYDRGPQPDRIEELINYSVDIQWGNHVDVLWIGAYSGSKVCL 271

Query: 241 VNVIRIAARYNNITLIEDRYGINLRRLVDYSRRYYEPLPSFVPILDGEEMTHPDELDDL 300
N+IRI ARY+N+ +IED YGINLR L++ + +YY+ P+F P D E DE+ +
Sbjct: 272 ANIIRICARYDNLDIIEDVYGINLRPLNLAEKYYDDNPAFRPKAD--ENRPEDEIKQIT 329

Query: 301 MIQQATAILOQFKLEAQLIDRRPEFQMHNRQLINQVNYKDLSSISKEVHVHQLKDFNSRCID 360
I QA A++QFKLE+ +I RRP F M R L+ +++Y I++ +QL++ I+
Sbjct: 330 KIHQAIAMIQFKLESPIIKRRPNFNMEERLLEKIDYDKNETITNGKTYQLENTCFATIN 389

-1992-

Query: 361 SKNPSRLTSEEEELLQQLMIAFQTSESLLKKHIDFLFEKGSMTLYNDNLLFHGCIPMHSN 420
 + P +L EE E++ +L+ + Q SE L +H++F+ +KGS+YL YN NLL HGCIP+ N
 Sbjct: 390 PEQPDQLLEEEAEVIDKLLFSVQHSEKLRHMFMMKGSLLYKYNGLLIHGCIPVDEN 449

Query: 421 GDFKSFKIAGKTYGGRDLLDLFESQIRLAYARPEKHDDLATDIWYLCWGENSSSLFGKNA 480
 G+ ++ I K Y GR+LLD+FE +R A+A PE+ DDLATD+ WYLW GE SSLFGK A
 Sbjct: 450 GNMETMMIEDKPYAGRELLDVFERFLREAFAPHEETDDLATDMAWYLTGEYSSSLFGKRA 509

Query: 481 MTTTFERYYYSDKVTQERKNPYFKLRDKDDICTALLQEFDL-PKFGHIVNGHTPVKEKNG 539
 MTTTFERY++ +K TH+E+KNPY+ LR+ + C +L EF L P GHI+NGHTPVKE G
 Sbjct: 510 MTTTFERYFIKEKETHKEKKNPYYYLREDEATCRNILAEFGLNPDHGHINGHTPVKEIEG 569

Query: 540 EQPIKANGKMLVIDGGFAKGYQKNTGLAGYTLIYNSYGIQLISHLPTSIEEVLSTGNYI 599
 E PIKANGKM+VIDGGF+K YQ TG+AGYTL+YNSYG+QL++H F S EVLS +
 Sbjct: 570 EDPIKANGKMIVIDGGFSKAYQSTTGLAGYTLIYNSYGMQLVAHKHFNKA EVLSTGTDV 629

Query: 600 IDTKRLVEEAKDRILVKDTTIGQKLTKKIDLDHL--YRHFQ 639
 + KRLV++ +R VK+T +G++L +E+ L+ L YR+ +
 Sbjct: 630 LTVKRLVDKELERKKVKETNVGEELLQEVAILSESLREYRYMK 671

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 5486 (GBS197) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 168 (lane 17 & 18; MW 89kDa) and in Figure 169 (lane 2; MW 89kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 6; MW 99kDa).

Purified Thio-GBS197-His is shown in Figure 244, lane 6.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1766

A DNA sequence (GBSx1873) was identified in *S.agalactiae* <SEQ ID 5487> which encodes the amino acid sequence <SEQ ID 5488>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2433(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12719 GB:Z99108 alternate gene name: ygaP~similar to
 hypothetical proteins [Bacillus subtilis]
 Identities = 176/367 (47%), Positives = 240/367 (64%), Gaps = 6/367 (1%)

Query: 3 IKAEIQKLAKKEIGISKIGFTTADNFDYLEKSLRASVEEGRNSGFHEHKVIEDRIYPERLLE 62
 +K E+ + AK IG+ KIGFTTAD FD L+ L G SGFE IE R+ P+ LL
 Sbjct: 55 LKEELIEYAKSIGVDKIGFTTADTFDSLKDRLILQESLGYLSGFEEPDIEKRVTPKLLLP 114

Query: 63 SAKTIISIGVAYPHKLPQQPQKT-SYKRGKITPNSWGLDYHYVVGEKLDRLSKGIEELCR 121
 AK+I++I +AYP ++ P+ T + +RG SWG DYH V+ EKLD L ++
 Sbjct: 115 KAKSIVAIALAYPSRMKDAPRSTRTERRGIFCRASWGKDYHDVLRKLDLLEDFLKSKE 174

Query: 122 DFPLQQKAMVDTGALVDTAVAQRAGIGFIGKNGLVISKEYGSYMFGLGELITNLEIEPDKP 181
 D ++ K+MVDTG L D AVA+RAGIGF KN ++ + EYGSY++L E+ITN+ EPD P
 Sbjct: 175 D--IRTKSMVDTGELSDRAVAERAGIGFSAKNCMITTPYGSYVLAEMITNIPFEPDVP 232

-1993-

Query: 182 VDYDCGDCRRCLDACPTSCSLIGDGSMAKRLCSFQTQDKGMMDIEFRKKIKTVIYGCDIC 241
 ++ CG C +CLDACPT L+ G +NA+RC+SF TQ KG + EFR KI +YGCD C
 Sbjct: 233 IEDMCGSCTKCLDACPTGALVNPQLNAQRCISFLTQTKGFLPDEFRTKIGNRLYGCDTC 292

Query: 242 QICCPYNGKINNPLATEI--DPELAQPELIPFLSLSNGQFKEKFGMIAGSWRGKNILQRN 299
 Q CP NKG + L E+ DPE+A+P L P L++SN +FKEKFG ++GSRGK +QRN
 Sbjct: 293 QTVCPNLNGKDFHLHPEMEPDPEIAKPLLKPLLAISNREFKEKFGHVSGSWRGKKPIQRN 352

Query: 300 AIIALANAHDKTAVVKLIEIIDKNNNPIHTATAIWALGEIVKKPNDEILEFMSNLTLDKDE 359
 AI+ALA+ D +A+ +L E++ K+ P+ TA WA+G+I E LE KDE
 Sbjct: 353 AIIALAHFKDASALPELTELHMKDPRPVIRGTAAWAIGKIGDPAYAELEKALEKE-KDE 411

Query: 360 DSRKELE 366
 +++ E+E
 Sbjct: 412 EAKLEIE 418

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5489> which encodes the amino acid sequence <SEQ ID 5490>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3337(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 363/374 (97%), Positives = 367/374 (98%)

Query: 1 MDIAEKIQLAKEIGISKIGFTTADNFDYLEKSLRASVEEGRNSGFEHKVIEDRIYPERL 60
 M IKA EI+ LAKEIGISKIGFTTADNFDYLEKSLRASVEEGRNSGFEHKVIEDRIY ERL
 Sbjct: 18 MTKAEIKALAKEIGISKIGFTTADNFDYLEKSLRASVEEGRNSGFEHKVIEDRIYTERL 77

Query: 61 LESAKTIISIGVAYPHKLPQQPQKTSYKRGKITPNSWGLDYHYVVGEKLDRLSKGIEELC 120
 LESAKTIISIGVAYPHKLPQQPQKT YKRGKITP+SWGLDYHYVVGEKLDRLSKGIEELC
 Sbjct: 78 LESAKTIISIGVAYPHKLPQQPQKTPYKRGKITPSSWGLDYHYVVGEKLDRLSKGIEELC 137

Query: 121 RDFPLQQKAMVDTGALVDTAVAQRAGIGFIGNGLVISKEYGSYMF LGELITNLEIEPDK 180
 RDFPLQQKAMVDTGALVDTAVAQRAGIGFIGNGLVISKEYGSYMF LGELITNLEIEPDK
 Sbjct: 138 RDFPLQQKAMVDTGALVDTAVAQRAGIGFIGNGLVISKEYGSYMF LGELITNLEIEPDK 197

Query: 181 PVDYDCGDCRRCLDACPTSCSLIGDGSMAKRLCSFQTQDKGMMDIEFRKKIKTVIYGCDI 240
 PVDYDCGDCRRCLDACPTSCSLIGDGSMAKRLCSFQTQDKGMMDIEFRKKIKTVIYGCDI
 Sbjct: 198 PVDYDCGDCRRCLDACPTSCSLIGDGSMAKRLCSFQTQDKGMMDIEFRKKIKTVIYGCDI 257

Query: 241 CQICCPYNGKINNPLATEIDPELAQPELIPFLSLSNGQFKEKFGMIAGSWRGKNILQRNA 300
 CQICCPYNGKINN ATEIDPELAQPELIPFLSLSNG+FKEKFGMIAGSWRGKNILQRNA
 Sbjct: 258 CQICCPYNGKINN SPATEIDPELAQPELIPFLSLSNGKFKEKFGMIAGSWRGKNILQRNA 317

Query: 301 IIALANAHDKTAVVKLIEIIDKNNNPIHTATAIWALGEIVKKPNDEILEFMSNLTLDKDED 360
 IIALANAHDKTAVVKLIEIIDKNNNPIHTATAIWALGEIVKKPNDEIL FMS+LITLDKDED
 Sbjct: 318 IIALANAHDKTAVVKLIEIIDKNNNPIHTATAIWALGEIVKKPNDEILAFMSHILTLDKDED 377

Query: 361 SRKELELIRHKWQF 374
 SRKELELIRHKWQF
 Sbjct: 378 SRKELELIRHKWQF 391

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1994-

Example 1767

A DNA sequence (GBSx1874) was identified in *S.agalactiae* <SEQ ID 5491> which encodes the amino acid sequence <SEQ ID 5492>. This protein is predicted to be peptide chain release factor 2 , fragment (prfB). Analysis of this protein sequence reveals the following:

```

5   Possible site: 23
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.4903(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAC67303 GB:AF017113 putative peptide chain release factor RF-2
      [Bacillus subtilis]
      Identities = 194/336 (57%), Positives = 251/336 (73%), Gaps = 2/336 (0%)

Query: 2   EEEIALLLENQMTEPDFWNDNIAAQKTSQELNELKGKYDTFHNMQELSDETELLLEMLDE- 60
Sbjct: 30  EARIAELDEQMADEPFWNDQQAQTVINEANGLKDYVNSYKKLINESHEELQMTDHLLEE 89

Query: 61  -DDSLKEELEENLMQLDKIMGAYEMTLLLSEPYDHNNAILIHPGSGGTEAQDWGDLLELR 119
Sbjct: 90  PDTDLQLELEKELKSLTKEFNEFELQLLLSEPYDKNNAILELHPGAGGTESQDWGSMLLR 149

Query: 120 MYTRFGNANGFKVEVLDYQAGDEAGIKSVTLSEFEGPNAYGLLKSEMGVHRLVRISEPFDSA 179
Sbjct: 150 MYTRWGERRGFKVETLDYLPGEAGIKSVTLLIKGHNAYGYLKAEGKVHRLVRISEPFSS 209

Query: 180 KRRHTSFASVEVMPPELDDTIEVEVRDDDIKMDTFRSGGAGGQNVNKVSTGVRLTHIPTGI 239
Sbjct: 210 GRRHTSFVSCVEMPEFNDEIDIDIRTEDIKVDTYRASGAGGQHVNTTDSAVRITHLPTNV 269

Query: 240 VVSSTVDRTQYGNRDRAMKMLQAKLYQLEQEKKAQEVDAKGDKEITWGSQIRSYVFTP 299
Sbjct: 270 VVTCQTERSQIKNRERAMKMLKAKLYQRRIEEQAELEIRGEQKEIGWSQIRSYVFHP 329

Query: 300 YTMVKDHRTNFELAQVDKVMDEINGFIDAYLKWRI 335
Sbjct: 330 YSMVKDHRTNTEMGNVQAVMDGDIDTFIDAYLRSKL 365
40

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5493> which encodes the amino acid sequence <SEQ ID 5494>. Analysis of this protein sequence reveals the following:

```

45   Possible site: 23
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.4779(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
50          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 334/337 (99%), Positives = 336/337 (99%)

55   Query: 1   MEEIALLLENQMTEPDFWNDNIAAQKTSQELNELKGKYDTFHNMQELSDETELLLEMLDE 60
      +EEIALLLEN MTEPDFWNDNIAAQKTSQELNELKGKYDTFHNMQELSDETELLLEMLDE
      Sbjct: 1   LEEIALLLENHMTTEPDFWNDNIAAQKTSQELNELKGKYDTFHNMQELSDETELLLEMLDE 60

Query: 61  DDSLKEELEENLMQLDKIMGAYEMTLLLSEPYDHNNAILIHPGSGGTEAQDWGDLLELR 120
Sbjct: 61  DDSLKEELEENLMQLDKIMGAYEMTLLLSEPYDHNNAILIHPGSGGTEAQDWGDLLELR 120
60
Sbjct: 61  DDSLKEELEENLMQLDKIMGAYEMTLLLSEPYDHNNAILIHPGSGGTEAQDWGDLLELR 120

```

-1995-

Query: 121 YTRFGNANGFKVEVLDYQAGDEAGIKSVTLSEFGPNAYGLLKSEMGVHRLVRISPFDSAK 180
 YTRFGNANGFK+EVLDYQAGDEAGIKSVTLSEFGPNAYGLLKSEMGVHRLVRISPFDSAK
 Sbjct: 121 YTRFGNANGFKIEVLDYQAGDEAGIKSVTLSEFGPNAYGLLKSEMGVHRLVRISPFDSAK 180

Query: 181 RRHTSFASVEVMEPELDDTIEVEVRDDDIKMDTFRSGGAGGQNVNKVSTGVRLTHIPTGIV 240
 RRHTSFASVEVMEPELDDTIEVEVRDDDIKMDTFRSGGAGGQNVNKVSTGVRLTHIPTGIV
 Sbjct: 181 RRHTSFASVEVMEPELDDTIEVEVRDDDIKMDTFRSGGAGGQNVNKVSTGVRLTHIPTGIV 240

Query: 241 VSSTVDRTQYGNRDRAMKMLQAKLYQLEQEKKAQEVDALKGDKEITWGSQIRSYVFTFY 300
 VSSTVDRTQYGNRDRAMKMLQAKLYQLEQEKKAQEVDALKGDKEITWGSQIRSYVFTFY
 Sbjct: 241 VSSTVDRTQYGNRDRAMKMLQAKLYQLEQEKKAQEVDALKGDKEITWGSQIRSYVFTFY 300

Query: 301 TMVKDHRITNFELAQVDKVMGEINGFIDAYLKWRIED 337
 TMVKDHRITNFELAQVDKVMGEINGFIDAYLKWRIED
 Sbjct: 301 TMVKDHRITNFELAQVDKVMGEINGFIDAYLKWRIED 337

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1768

A DNA sequence (GBSx1875) was identified in *S.agalactiae* <SEQ ID 5495> which encodes the amino acid sequence <SEQ ID 5496>. This protein is predicted to be cell-division ATP-binding protein (ftsE). Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3928(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC67262 GB:AF017113 cell division ATP-binding protein [Bacillus subtilis]
 Identities = 138/228 (60%), Positives = 179/228 (77%)

Query: 3 LIEMSGVTKKYRRSTTALRNLSIQQGEFVYLVGPSGAGKSSLIRLLYREEKLSSGRLK 62
 +IEM V K Y .AL ++++I GEFVY+VGPSGAGKS+ I+++YREEK + G++
 Sbjct: 1 MIEMKEVYKAYPNGVKALNGISVTIHPGEFVYVVGPSGAGKSTFIKMIYREEKPTKGQIL 60

Query: 63 VGEFNLNKLKRRQIPILRRSIGVVFQDYKLLPTKTVYENAVAFAMQVIGAKRRHIKRVPE 122
 + +L +K ++IP +RR IGVVFQD+KLLP TV+ENVAFA++VIG + IKRV E
 Sbjct: 61 INHKDLATIKEKEIPFVRRKIGVVFQDFKLLPKLTVFENVAFALEVIGEOPSVIKRVLE 120

Query: 123 VLELVGLKHKMRSFPTQLSGGEQQRVAIARAIVNNPKLLIADEPTGNLDPETIAWEIMHLL 182
 VL+LV LKHK R FP QLSGGEQQRV+IAR+IVNNP ++IADEPTGNLDP+ +WE+M L
 Sbjct: 121 VLDLVQLKHKARQFPDQLSGGEQQRVSIARSIVNNPDVVIADPTGNLDPDTSWEVMKTL 180

Query: 183 ERINLQGTTVLMATHNSQIVNTLRHRVIEIEAGSVIRDEEKGEYGYHD 230
 E IN +GTTV+MATHN +IVNT++ RVI IE G ++RDE +GEYG +D
 Sbjct: 181 EEINNRRGTTVMATHNKEIVNTMKRRVIAIEDGIIVRDESRGEYGSYD 228

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5497> which encodes the amino acid sequence <SEQ ID 5498>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3728(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1996-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 191/230 (83%), Positives = 214/230 (93%)

5
Query: 1 MALIEMSGVTKKYRRSTTALRNLNLSIQGEFVYLVGPSGAGKSSLIRLLYREEKLSSGR 60
MALIEMSGVTKKYRRSTTALR++N+S+ QGEFVYLVGPSGAGKS+ I+LLYREE+L++G+
Sbjct: 1 MALIEMSGVTKKYRRSTTALRDVNVSNQGEFVYLVGPSGAGKSTFIKLLYREEQLTTGK 60

10
Query: 61 LKVGEFNLNKLKRRQIPILRRSIGVVFQDYKLLPTKTVYENVAFAMQVIGAKRRHIKKRV 120
L VGEFNL KLK R +PILRR IGVVFQDYKLLP KTV+ENVA+AM+VIG KRRHIKKRV
Sbjct: 61 LYVGEFNLTKLKARDVPILRRHIGVVFQDYKLLPRKTVFENVAYAMEVIGEKRRHIKKRV 120

15
Query: 121 PEVLELVGLKHKMRSFPTQLSGGEQQRVAIARAIVNNPKLLIADEPTGNLDPEIAWEIMH 180
PEVL+LVGLKHKMRSFP+QLSGGEQQRVAIARAIVNNPKLLIADEPTGNLDPEI+WEIM
Sbjct: 121 PEVLDLVGLKHKMRSFPSQLSGGEQQRVAIARAIVNNPKLLIADEPTGNLDPEISWEIMQ 180

20
Query: 181 LLERINLQGTTVLMATHNSQIVNTRLRHRVIEIEAGSVIRDEEKGEGYGYHD 230
LLERIN+QGT+LMATHNS IVNT RHRV+ IE G ++RDEEKG+YGY D
Sbjct: 181 LLERINVQGTTLIMATHNSHIVNTRFHRVVAIEDGRIVRDEEKGEGYGYDD 230

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1769

25 A DNA sequence (GBSx1876) was identified in *S.agalactiae* <SEQ ID 5499> which encodes the amino acid sequence <SEQ ID 5500>. This protein is predicted to be ftsE protein (ftsX). Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have no N-terminal signal sequence

30
INTEGRAL Likelihood = -10.77 Transmembrane 296 - 312 (291 - 322)
INTEGRAL Likelihood = -9.24 Transmembrane 203 - 219 (198 - 228)
INTEGRAL Likelihood = -6.16 Transmembrane 49 - 65 (40 - 68)
INTEGRAL Likelihood = -3.40 Transmembrane 255 - 271 (252 - 273)

35 ----- Final Results -----

bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9629> which encodes amino acid sequence <SEQ ID 9630> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC67264 GB:AF017113 cell division protein [Bacillus subtilis]

Identities = 112/311 (36%), Positives = 182/311 (58%), Gaps = 31/311 (9%)

45
Query: 27 RHFVESLKNLKRNFWMTFASVTSVTITLLVLGLFSSVLLNVEKLTIDVSGNFTISAFNLV 86
RH ES K+L RN WMTFAS+++VT+TL+LVG+F ++LN+ + T+ I +++
Sbjct: 7 RHLRESFKSLGRNTWMTFASISAVTVTLLVLGVFLVIMLNLNNMATNAEKQVEIKVLIDL 66

50
Query: 87 DSTDAQKQVKDKDKGLKDNPDYHKVYDKIKRISGVEKVITYSSKAEQLKEVQKEYGSDVID 146
+ D K +D K+ + IK + G++ VT+SSK ++L ++ +G
Sbjct: 67 TA-----DQKAQD-----KLQNDIKELKGIQSVTFSSKEKELDQLVDSFGDSGKS 111

55
Query: 147 DTYKDA---LLDVYVVGTSKAKSVSEAGRIEGRV---DYTKPIDST-KLSNLTNDNI 199
T KD L D +VV T+ + +V++ I +++ V Y KE + K+ ++ NI
Sbjct: 112 LTMKDQENPLNDAFVVKTTDPHDTPNVAKKIEKMDHVYKVITYGKEEVSRLEFKVVGVSRI 171

Query: 200 RIWFGGGVALLIVL---AIFLISNTIRMSIMSRRTDIEIMRLVGAKNSYIRGPFFFEGAW 256

-1997-

G+AL+I L A+FLISNTI+++I +RR +IEIM+LVGA N +IR PFF EG
 Sbjct: 172 -----GIALIIGLVFTAMFLISNTIKITIFARRKEIEIMKLVGATNWFIRWPFLEGLL 225

Query: 257 VGILGAIVPSLIFYFGYQFVNKFNPKFETSHVSLYPMDIMVPAIIGGMVIIGIIGSLG 316
 +G+ G+++P + YQ+V PK+ S VSL P + V + ++ IG +IG G
 Sbjct: 226 LGVFGSVIPIALVLSTYQYVIGWVVPKVQGSFVSLLPYNPFVQVSLVLIAIGAVIGVWG 285

Query: 317 SVLSMRRYLKI 327
 S+ S+R++L++
 Sbjct: 286 SLTSIRKFLRV 296

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5501> which encodes the amino acid sequence <SEQ ID 5502>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -7.70	Transmembrane	195 - 211 (189 - 219)
INTEGRAL	Likelihood = -6.74	Transmembrane	39 - 55 (30 - 58)
INTEGRAL	Likelihood = -5.52	Transmembrane	294 - 310 (288 - 314)
INTEGRAL	Likelihood = -1.49	Transmembrane	246 - 262 (245 - 263)

----- Final Results -----
 bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC67264 GB:AF017113 cell division protein [Bacillus subtilis]
 Identities = 117/311 (37%), Positives = 184/311 (58%), Gaps = 19/311 (6%)

Query: 11 MIRYFFRHIWESIKNLKRNFWMTFASVSMVAVTLLVGVFAATLLNIQRVASGVENNVHI 70
 MI+ RH+ ES K+L RN WMTFAS+S V VTL LVGVF +LN+ +A+ E V I
 Sbjct: 1 MIKILGRHLRESFKSLGRNTWMTFASISAVTVTLILVGVFLVIMLNLNMATNAEKQVEI 60

Query: 71 NTYLQVDSTDAKVIQNTAGEPVNNDNYHSVYDKIAQIKGVKKITFSSKDEQLKKLOETL 130
 + + + A+ + + ND I ++KG++ +TFSSK+++L +L ++
 Sbjct: 61 KVLIDLTDQKAQ-----DKLQND-----IKELKGIQSVTFSSKEKELDQLVDSE 105

Query: 131 GDVWN---MYDQDTNPLQDIYLIETQTPKQVKAITKKIRTIEGVEAADYGGINSKDLFKF 187
 GD M DQ+ NPL D +++++T P + KKI ++ V YG +LFK
 Sbjct: 106 GDSGKSLTMKDQE-NPLNDAFVVKTTDPHDTNPAKKIEKMDHVYKVITYGKEEVSRLLFKV 164

Query: 188 STLIQTWGLIGTAMLLFVAVFLISNTIRMTIMSRKRDIEMRLVGAKNSYIRGPFFFEGA 247
 + + G+ L+F A+FLISNTI++TI +R+++IEIM+LVGA N +IR PFF EG
 Sbjct: 165 VGVSRNIGIALIIGLVFTAMFLISNTIKITIFARRKEIEIMKLVGATNWFIRWPFLEGL 224

Query: 248 WVGLLGAIVPSLLIYYGYDLVYKHFAQELQRNNLSMYPLDPYVYYLIGALFVIGIMIGSL 307
 +G+ G+V+P L+ Y V ++Q + +S+ P +P+V+ + L IG +IG
 Sbjct: 225 LLGVFGSVIPIALVLSTYQYVIGWVVPKVQGSFVSLLPYNPFVQVSLVLIAIGAVIGVW 284

Query: 308 GSVLSMRRYLK 318
 GS+ S+R++L+
 Sbjct: 285 GSLSIRKFLR 295

An alignment of the GAS and GBS proteins is shown below.

Identities = 173/318 (54%), Positives = 238/318 (74%), Gaps = 5/318 (1%)

Query: 13 MKRRENMVIMIN-FFRHFWESLKNLKRNFWMTFASVTSVTITLLLVGLFSSVLLNVEKLT 71
 MK++E MV MI FFRH WES+KNLKRNFWMTFASV+ V +TL LVG+F++ LLN++++
 Sbjct: 2 MKKKEIMVTMIRYFFRHIWESIKNLKRNFWMTFASVSMVAVTLLVGVFAATLLNIQVA 61

Query: 72 TDVSGNFTISAFNLVDSTDAKQKQVKDKGKLDNPDYHKVYDKIKRISGVEKVITYSSKAE 131
 + V N I+ +L VDSTDA K +++ G+ +N +YH VYDKI +I GV+K+T+SSK E
 Sbjct: 62 SGVENNVHINTYLQVDSTDAKVIQNTAGEPVNNDNYHSVYDKIAQIKGVKKITFSSKDE 121

-1998-

Query: 132 QLKEVQKEYGSDVID--DTYKDALLDVYVGTSSAKVSKSVSEAIGRIEGVDYTKEP-ID 188
 QLK++Q+ G DV + D + L D+Y++ T + K K++++ I IEGV+ I+
 Sbjct: 122 QLKKLQETLG-DVWNMYDQDTNPLQDIYLIETQTPKQVKATKKIRTIEGVAAADYGGIN 180

5 Query: 189 STKLSNLTDNIRIWGFGGVALLIVLAIFLISNTIRMSIMSRRTDIEIMRLVGAKNSYIRG 248
 S KL + I+ WG G A+L+ +A+FLISNTIRM+IMSR+ DIEIMRLVGAKNSYIRG
 Sbjct: 181 SDKLKFSTLIQTWGLIGTAMLLFVAVFLISNTIRMTIMSRKRDIEIMRLVGAKNSYIRG 240

10 Query: 249 PFFFEGAWVGILGAIVPSLIFYFGYQFVFNKFNPKFETSHVSLYPMDIMVPAIIGGMVII 308
 PFFFEGAWVG+LGA++PSL+ Y+GY V+ F + + +++S+YP+D V +IG + +I
 Sbjct: 241 PFFFEGAWVGLLGAVLPSLLIYYGYDLVYKHFAQELQRNNLSMYPLDPYVYYLIGALFVI 300

Query: 309 GIIIGSLGSLVLSMRRYLK 326
 GI+IGSLGSLVLSMRRYLK
 15 Sbjct: 301 GIMIGSLGSLVLSMRRYLK 318

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1770

20 A DNA sequence (GBSx1877) was identified in *S.agalactiae* <SEQ ID 5503> which encodes the amino acid sequence <SEQ ID 5504>. This protein is predicted to be carboxymethylenebutenolidase-related protein. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have a cleavable N-term signal seq.

25

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF10898 GB:AE001979 carboxymethylenebutenolidase-related
 protein [Deinococcus radiodurans]

Identities = 65/183 (35%), Positives = 98/183 (53%), Gaps = 3/183 (1%)

35

Query: 56 SKGKVKANIIFYQGALVEEEAYSQ LARDLADKGDNTYILKTPLNLPVLSPHKAKTIINQN 115

+ +VK ++FY G V +AY L R LA +G T I PL+L + +A+ +I +

Sbjct: 100 ASAEVKTLLVFYPPGGRVPRQAYEWLGRALAVRGVQTVIPAPFLDLAITGTERAEGLIARY 159

40

Query: 116 HL-TNVYLAGHSLGGVVASQNAKVAP--VRGLILLASYP SRKSDLSHKNLRVLSITASND 172

V LAGHSLGG VA+Q A + P + GL+LLA+YP+ +L LS+ A D

Sbjct: 160 GAGKRVVLAGHSLGGTVAAQYAALRPDKIDGLLLLAAYPAPNVNLHDARFPALSLLAEKD 219

45

Query: 173 HILNWEKYEEAKRLPNSSFTFTIVGGNHSRFGNYGHQKGDGKATLSHKSSEKQLATFIS 232

+ + +RLP ++ + G HS FG YG Q+GDG T+S +E+++ +

Sbjct: 220 GVADAGLVRGGLERLPKNTRLTLVLPAGVHSFFGRYGPQQGDGVPTVSRARAEREIVQAVE 279

Query: 233 NFI 235

FI

50

Sbjct: 280 TFI 282

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 5504 (GBS158) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 4; MW 27kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 5; MW 52kDa).

55

-1999-

The GBS158-GST fusion product was purified (Figure 113; see also Figure 201, lane 4) and used to immunise mice (lane 1+2 product; 14.5µg/mouse). The resulting antiserum was used for Western blot, FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1771

A DNA sequence (GBSx1878) was identified in *S.agalactiae* <SEQ ID 5505> which encodes the amino acid sequence <SEQ ID 5506>. Analysis of this protein sequence reveals the following:

10 Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.0281(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:BA06539 GB:AP001516 unknown conserved protein [Bacillus halodurans]
Identities = 83/197 (42%), Positives = 114/197 (57%), Gaps = 4/197 (2%)

Query: 35 NTYYLVNDQAV-ILIDPGSNGQEIIAKIKSFEKPLVAILLTHTHYDHIFSLDLVRDTFDN 93
N Y NDQ I+ DPG +++I ++ + +AILLTH H+DHI +++ VR+TF +
Sbjct: 14 NNYIQTNQDQEGEIIIFDPGGEVEKLITWLRDRQITPLAILLTHAHFDHIGAVEDVRNTF-H 72

25 Query: 94 PPVYVSEKEAAWLSSPDDNLSGLGRHDDIINVIARPAENFFKLKQPYQLNGFEFTVLPTP 153
PVY+ E E WL P N S L I AR AE+ +Q + F + VL TP
Sbjct: 73 IPVYIHENEKEWLIDPQRNGSSLFIPGSSIK--AREAEHLITGEQDLSIGSFSYQVLETP 130

30 Query: 154 GHSWGGVSFVFHSDELVTGDALFRETIGRTDLPTSNFEDLITGIRQELFTLPSHYSVHP 213
GHS G +S+ D++V +GDALF +IGRTDLP + + L+ I +L LP +V
Sbjct: 131 GHSPGSLSYAKEDKIVFSGDALFAGSIGRTDLPGGDHQLLLDSIHDKLELPEDTTVAS 190

35 Query: 214 GHGMNTTIGHEKNFNP 230
GHG TTIGHE + NPF
Sbjct: 191 GHGPTTTIGHEMDGNPF 207

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5507> which encodes the amino acid sequence <SEQ ID 5508>. Analysis of this protein sequence reveals the following:

40 Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.0407(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 217/231 (93%), Positives = 224/231 (96%)

50 Query: 1 MPFIFRHSFFNKVLIFWYTIIMKIYKTINHIAGENTYYLVNDQAVILIDPGSNGQEIIAK 60
+PFIFR+SFFNKVLIFWYTI+MKIYKTINHIAGENTYYLVNDQAVILIDPGSNGQEIIAK
Sbjct: 1 LPFIFRYSFFNKVLIFWYTIIMKIYKTINHIAGENTYYLVNDQAVILIDPGSNGQEIIAK 60

55 Query: 61 IKSFEKPLVAILLTHTHYDHIFSLDLVRDTFDNPPVYVSEKEAAWLSSPDDNLSGLGRHD 120
IKSFEKPLVAILLTHTHYDHIFSLDLVRD FD+PPVYVSEKEAAWLSSPDDNLSGLGRHD

-2000-

Sbjct: 61 IKSFEKPLVAILLTHYDHFISLDLVRDAFDHPPVYVSEKEAOWLSSPDDNLSGLGRHD 120

Query: 121 DIINVIARPAENFFKLKQPYQLNGFEFTVLPTPGHSWGGVSFVFHSDELVVTGDALFRET 180
DII VIARPAENFFKLKQPYQLNGFEFTVLPT GHSWGGVSFVFHSDELVVTGDALFRET

5 Sbjct: 121 DIITVIARPAENFFKLKQPYQLNGFEFTVLPTSGHSWGGVSFVFHSDELVVTGDALFRET 180

Query: 181 IGRDLPSTNFEDLITGIRQELFTLP SHYSVHPGHGMNTTIGHEKNFNPF 231
IGRDLPTSNFEDLITGIRQELFTLP+HY V+PGHG +TTI HEKN NPFF

10 Sbjct: 181 IGRDLPSTNFEDLITGIRQELFTLPNHYRVYPGHGPSTTICHEKNANPF 231

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1772

A DNA sequence (GBSx1879) was identified in *S.agalactiae* <SEQ ID 5509> which encodes the amino acid sequence <SEQ ID 5510>. This protein is predicted to be acetoin reductase (fabG). Analysis of this protein sequence reveals the following:

Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1596(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9631> which encodes amino acid sequence <SEQ ID 9632> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC48769 GB:U71200 acetoin reductase [Bos taurus]
Identities = 162/254 (63%), Positives = 188/254 (73%), Gaps = 2/254 (0%)

Query: 12 KVAIVTGAGQGIGFALAKRLHADGFKIGVLDYNEETAQAQAVDKLSPED--AVAVVADVSK 69
KVA+VTG QGIG AI L ADGF + V D NE ++ + A+AV DVS

Sbjct: 4 KQAMVTGGAQGIGEAIVXXLSADGFAVAVADLNEAKSKXVATDIEKNGGTALAVKLDVSD 63

Query: 70 RDQVFDAFQKVVDTFGDLNVVNNAGVAPTTPLDTITEEQFEKAFAINVGGTIWGSQAAQ 129
R+ F A ++V + G +V+VNNAG+ PTP+DTIT E F+K + INV G IWG QAA

Sbjct: 64 REGFFAAVKEVAEKLGGFDVLVNNAGLGPTTPIDTITPELFDKVYHINVAGDIWGIQAAV 123

Query: 130 KHRELGHGGKIINATSQAGCEGNPNLTVYGGTKFAVRGITQTLAKDLASEGITVNAYAP 189
+ F++ G+GGKIINATSQAG GNPNL++Y TKFAVR +T A+DLA + ITVNAYAP

Sbjct: 124 EQFKKNGNGGKIINATSQAGVGNPNLSLYSSTKFAVRCLTPVAARDLAEQNITVNAYAP 183

Query: 190 GIVKTPMMFDIAHEVGKNAGKDEWGMQFAKDITLKRLEPEDVANAVGFLAGDSDNYI 249
GIVKTP FDIAHEVGKNAGKDEWGM+ FAKDI LKRLSEPEDVA AV FLAG DSDNYI

45 Sbjct: 184 GIVKTPXXFDIAHEVGKNAGKDEWGMQTFKDI LKRLSEPEDVAAVAFLAGDSDNYI 243

Query: 250 TGQTIIVVDGGMVVFH 263
TGQTI VDGGM FH

50 Sbjct: 244 TGQTIEVDGGMQFH 257

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5511> which encodes the amino acid sequence <SEQ ID 5512>. Analysis of this protein sequence reveals the following:

Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1131(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>